

Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia
Departamento de Engenharia Electrotécnica

*Desenho e implementação de um sistema computacional para apoiar a
gestão de projectos utilizando técnicas de data mining*

Por
Vitor Miguel Marques Parada

Dissertação apresentada na Faculdade de Ciências e Tecnologia da
Universidade Nova de Lisboa para obtenção do grau de Mestre em
Engenharia Electrotécnica e de Computadores

Orientador: Professor Celson Lima

Lisboa
2010

Agradecimentos

Gostaria de agradecer a todas as pessoas que directa ou indirectamente contribuíram para a concretização deste projecto/dissertação.

Ao Professor Celson Lima, meu orientador, pelo sentido crítico que me soube inculcar e pela partilha constante de ideias e conhecimento que me ajudaram a desenvolver o presente trabalho.

Aos membros do grupo de trabalho, destacando a colaboração do João Antunes.

Também não posso deixar de agradecer a disponibilidade do João Pinto, em vários momentos complicados de programação.

Gostaria também de agradecer a toda a minha família, nomeadamente aos meus pais pelo apoio incondicional e por estarem sempre presentes.

Por último, agradeço à Ana, o apoio, a paciência, a dedicação, a presença, enfim um sem número de atributos, os quais não consigo descrever por palavras. Sem ti a realização desta dissertação e respectivo mestrado não seria possível.

Muito obrigado a todos.

Resumo

A colaboração tem sido uma mais-valia nas empresas baseadas em conhecimento pois permite a partilha de conhecimento entre funcionários, proporcionando melhorias ao nível do seu desempenho. De facto, os funcionários que utilizam as informações partilhadas podem elaborar estratégias mais eficazes e criativas, o que se traduz em vantagens competitivas.

Neste sentido, estas empresas têm suportado a ideia da criação de ambientes de trabalho colaborativos capazes de estimular os indivíduos a partilhar conhecimento. Os ambientes de trabalho colaborativos visam não só promover a partilha de conhecimento mas também aumentar o conhecimento existente, nomeadamente pela descoberta de novo conhecimento a partir de dados armazenados.

O rápido avanço das Tecnologias de Informação deu um grande contributo à área de descoberta de conhecimento em bases de dados, pois permitiu o armazenamento de quantidades consideráveis de dados em computadores. Pelo facto da análise humana do conteúdo destes dados ser limitada, falhando na detecção de detalhes importantes, foram desenvolvidas ferramentas de *data mining* para descoberta automática de nova informação a partir de dados armazenados.

Assim, o propósito deste trabalho é desenhar e implementar uma infra-estrutura de *software* que permita descobrir, a partir de bases de dados, informação não acessível pelos métodos tradicionais de análise, recorrendo para isso a conceitos e técnicas de *data mining*.

A infra-estrutura de *software* desenvolvido, *Miner*, foi integrado e testado no contexto do Projecto Europeu CoSpaces e permitiu a obtenção de informações importantes sobre projectos e que podem ser fundamentais durante o processo de tomada de decisão.

Abstract

Collaboration has been seen as an asset in knowledge based-organizations allowing the share of knowledge among employees and providing them better performances. Indeed, the employees are able to use the collected information in order to develop effective and creative strategies yielding competitive advantages.

Accordingly, the creation of collaborative workplaces to stimulate employees to share knowledge has been supported by these organizations. Collaborative workplaces are intended not only to promote the sharing of knowledge but also to improve the existing knowledge, namely the discovery of knowledge from stored data.

The quick progress on information technologies was fundamental to the field of knowledge discovery in databases as it allowed the storage of large amounts of data in computers. Given the fact that the human analysis of this data is limited, failing to detect important details, data mining tools were developed in order to allow automatic discovery of new information using stored data.

Thus, the main purpose of this study is to design and implement a software infrastructure to help of discovering new information from databases using data mining tools.

The software infrastructure developed here, called *Miner*, was integrated and tested in the European research project CoSpaces and provided important information about projects that may be fundamental during the decision making process.

Lista de acrónimos

API - Application Programming Interface

BSCW - Basic Support for Cooperative Work

CoSKS - CoSpaces Knowledge Support

CSF - CoSpaces Software Framework

DER – Diagrama de Entidade Relação

DM - Data Mining

HTML - HyperText Markup Language

ICE - Interface Controlo e Entidade

IDE - Integrated Development Environment

JDBC - Java Database Connectivity

JSP - Java Server Pages

MVC - Model View Controller

ORM - Object Relational Mapping

SQL - Structured Query Language

TM - Text Mining

UML - Unified Modeling Language

XML - Extensible Markup Language

Lista de termos e expressões da língua inglesa

Case Based Reasoning

Clustering

Data mining

Framework

Fuzzy

Groupware

Knowledge discovery in databases

Marketing

Project post-mortem

Ranking

Similar Time Series

Software

Stemming

Stop words

Text mining

User-friendly

Índice geral

1. Introdução	1
1.1 Objectivo	2
1.2 Contexto de desenvolvimento do trabalho.	2
1.3 Estrutura.....	3
2. Estado da arte	5
2.1 Gestão do conhecimento.....	5
2.1.1 Definição de gestão do conhecimento	5
2.1.2 Tipos de conhecimento	6
2.1.3 O conhecimento como recurso das empresas	9
2.2 Gestão do conhecimento colaborativa	11
2.2.1 O conceito de mercados de conhecimento	11
2.2.2 Tecnologias que suportam a colaboração em projectos de gestão do conhecimento.....	13
2.2.3 Ineficiências da gestão do conhecimento colaborativa	15
2.3 DM como ferramenta no processo de gestão do conhecimento	16
2.3.1 Etapas do processo de descoberta de conhecimento.....	16
2.3.2 Tarefas de DM	17
2.3.3 Técnicas de DM	19
2.3.4 Aplicações do DM	21
2.3.5 Text Mining.....	22
2.3.6 Análise de ferramentas de DM.....	23
2.4 Síntese	25
3. Requisitos e modelo conceptual	27
3.1 Termos	27
3.2 Os requisitos do projecto CoSpaces	29
3.3 Bases Conceptuais	33
3.4 DM como suporte à tomada de decisão	36
3.5 Fundamentos técnicos do CoSKS.....	38
3.6 O <i>Miner</i>	40
3.6.1 As interfaces do <i>Miner</i>	41

3.6.2 Os serviços do <i>Miner</i>	42
4. Implementação	47
4.1 Tecnologias adoptadas no trabalho	47
4.2 A infra-estrutura do software <i>Miner</i>	50
4.2.1 Formalização do <i>Miner</i>	51
4.2.2 Visão estrutural do <i>Miner</i>	56
4.2.2.1 O nível interface	57
4.2.2.2 O nível controlo	59
4.2.2.3 O nível entidade	61
4.2.2.4 Correspondência entre serviços <i>Miner</i> , métodos e tabelas.....	65
4.3 Descrição do contributo das ferramentas mais importantes na implementação do <i>Miner</i>	67
4.3.1 Utilização do BSCW	67
4.3.2 Utilização do RapidMiner.....	68
4.4 Pequeno exemplo de validação do sistema.....	70
5. Conclusões.....	77
5.1 Síntese do trabalho.....	77
5.2 Contribuição da pesquisa	78
5.3 Trabalhos futuros	79
Referências bibliográficas.....	81
ANEXO A - Diagrama de entidade relação	87
ANEXO B - BSCW.....	88
ANEXO C - Exemplos da utilização do <i>Miner</i>	89

Índice de figuras

Figura 2.1 - Quatro modos de conversão do conhecimento	8
Figura 2.2 - A espiral do conhecimento de Nonaka e Takeuchi	9
Figura 2.3 - Modelo para mercado do conhecimento proposto por Simard	12
Figura 2.4 - Categorização de ferramentas para conhecimento	14
Figura 2.5 - <i>Data mining</i> e o processo <i>knowledge discovery in database</i>	17
Figura 3.1 - Visão geral da arquitectura do CSF	30
Figura 3.2 - Requisitos do CoSKS	31
Figura 3.3 - Bases conceptuais do trabalho	33
Figura 3.4 - Ponto de decisões do projecto	34
Figura 3.5 - Níveis do conhecimento CoSpaces	35
Figura 3.6 - Fases do processo de colaboração	37
Figura 3.7 - Níveis técnicos do CoSKS	38
Figura 3.8 - Interfaces do <i>Miner</i>	41
Figura 3.9 - Serviços do <i>Miner</i>	43
Figura 4.1 - Interacções entre as tecnologias adoptadas	50
Figura 4.2 - Diagrama de casos de uso dos projectos	52
Figura 4.3 - Diagrama de casos de uso das reuniões	53
Figura 4.4 - Diagrama de casos de uso das tarefas	54
Figura 4.5 - Diagrama de casos de uso das questões	55
Figura 4.6 - Diagrama de casos de uso dos documentos	55
Figura 4.7 - Diagrama de classes ICE	57
Figura 4.8 - Interface projecto	57
Figura 4.9 - Interface reunião	58
Figura 4.10 - Interface questão	58
Figura 4.11 - Interface tarefa	58

Figura 4.12 - Interface documento	59
Figura 4.13 - Controlo projecto.....	60
Figura 4.14 - Controlo reunião.....	60
Figura 4.15 - Controlo questão.....	60
Figura 4.16 - Controlo tarefa	60
Figura 4.17 - Controlo documento	60
Figura 4.18 - Entidade projecto	61
Figura 4.19 - Excerto do DER (Ligações dos projectos)	62
Figura 4.20 - Entidade reunião.....	62
Figura 4.21 - Excerto do DER (Ligações das reuniões)	63
Figura 4.22 - Entidade questão	63
Figura 4.23 - Entidade tarefa.....	63
Figura 4.24 - Excerto do DER (Ligações das tarefas e questões)	64
Figura 4.25 - Entidade documento.....	64
Figura 4.26 - Excerto do DER (Ligações dos documentos).....	65
Figura 4.27 - Estrutura do BSCW	67
Figura 4.28 - Árvore de decisão obtida no RapidMiner	69
Figura 4.29 - Página de acesso à interface CoSKS	70
Figura 4.30 - Página inicial da interface CoSKS	71
Figura 4.31 - Resultados da utilização do serviço prever.....	72
Figura 4.32 - Resultados da utilização do serviço seleccionar actores	73
Figura 4.33 - Resultados da utilização dos serviços caracterizar e classificar	74

Índice de tabelas

Tabela 2.1 - Dois tipos de conhecimento	7
Tabela 2.2 - Ferramentas DM.....	25
Tabela 3.1 - Termos básicos e definições	27
Tabela 4.1 - Tecnologias adoptadas.....	47
Tabela 4.2 - Resumo dos serviços <i>Miner</i>	65

1. Introdução

A gestão do conhecimento tem vindo a impor-se como uma área em crescimento em vários domínios. Em grande parte, este crescimento deve-se à consciencialização de que os dados armazenados pelas empresas têm de ser necessariamente valorizados. Actualmente, o conhecimento constitui um componente crítico para o sucesso de uma empresa. Neste sentido, o principal objectivo da gestão do conhecimento é melhorar a performance das empresas, permitindo que os indivíduos, no seio dessas empresas, possam capturar, partilhar e aplicar o seu conhecimento para tomar decisões, ao mesmo tempo em que promovem inovações em termos de processos, produtos, estratégias de negócio, etc..

Por outro lado, o facto de as empresas estarem cada vez mais distribuídas geograficamente, obriga os funcionários a trabalharem com colegas distantes e esta foi uma das razões responsáveis pelo desenvolvimento de sistemas de colaboração. Estes sistemas são concebidos para melhorar os processos de tomadas de decisão em grupo, para facilitar a partilha de conhecimento e estimular a interacção entre os indivíduos, reduzindo as actividades desorganizadas. Deste modo, a gestão do conhecimento é inerentemente colaborativa e, na realidade, o que pretende é somar esforços e competências no sentido de atingir uma determinada meta que seria muito difícil de atingir individualmente.

O desenvolvimento de novas metodologias colaborativas permitirá criar um ambiente em que os indivíduos possam expressar a sua criatividade, o que constitui sem dúvida uma mais-valia para a gestão do conhecimento e logo para a organização e sucesso das empresas.

Em suma, as tecnologias de informação podem melhorar a organização das empresas por promover um conjunto de ferramentas colaborativas para a gestão do conhecimento.

1.1 Objectivo

No presente trabalho pretende-se desenvolver uma ferramenta de *software* capaz de melhorar o processo de gestão de projectos, baseando-se em conceitos e técnicas de *data mining* (DM), que permita capitalizar o conhecimento histórico criado em projectos semelhantes.

Esta ferramenta deve fornecer informações sobre casos históricos de projectos anteriores que facilitem o processo de tomada de decisões, nomeadamente:

- Oferecer um conjunto de soluções para auxiliar a resolução de um problema particular ou geral;
- Seleccionar participantes para uma tarefa ou projecto específico, de acordo com as suas características;
- Facultar informação relativa a custos e tempo de execução médios de projectos de uma determinada área; e
- Disponibilizar um conjunto de documentos relacionados com determinado assunto.

1.2 Contexto de desenvolvimento do trabalho.

Este trabalho foi desenvolvido no âmbito do projecto CoSpaces, um projecto integrado financiado pelo programa de Tecnologias de Informação da Sociedade (IST – Information Society Technologies) da Comissão europeia (EC-IST FP6), que pretende dar resposta às necessidades colaborativas, actualmente sentidas pelas grandes indústrias.

Na prática, o CoSpaces pretende desenvolver modelos de organizações e tecnologias que permitam suportar locais de trabalho colaborativos e inovadores, para equipas e indivíduos ligadas a empresas industriais distribuídas. A ideia é permitir parcerias para colaborações, melhorar a criatividade, aumentar a

produtividade e reduzir a duração dos ciclos de produção. Assim, o sucesso deste projecto dependerá inevitavelmente de uma actuação a vários níveis e que deverá incluir uma melhoria das técnicas para comunicação humana, o desenvolvimento de processos de visualização inovadores e um maior apoio ao conhecimento assim como interacção natural.

O componente do projecto CoSpaces que se relaciona directamente com o presente trabalho é o componente de conhecimento do CoSpaces (CoSpaces Knowledge Component). Este componente é constituído por um conjunto de serviços semânticos, direccionados para o conhecimento (CoSKS). Por sua vez, estes serviços devem possuir funcionalidades para apoiar as necessidades de conhecimento identificadas no projecto CoSpaces.

Este trabalho usa o modelo conceptual proposto para o CoSKS e explora a implementação de um novo módulo, responsável pela realização de *data mining*, de modo a contribuir para um melhor aproveitamento de conhecimento, no seio do modelo proposto pelo projecto CoSpaces.

1.3 Estrutura

O presente documento encontra-se organizado em cinco capítulos, conforme se descreve em seguida.

O Capítulo II aborda os fundamentos teóricos que suportam o tema do trabalho. Assim, descrevem-se algumas das definições mais importantes no contexto da gestão do conhecimento e dos processos de colaboração. Por outro lado, apresenta-se também uma descrição das principais técnicas e aplicações de DM.

No Capítulo III analisam-se os requisitos do CoSKS necessários para a implementação e interacção, do elemento responsável pela realização de processo de DM (*Miner*). Serão também descritos os serviços que o *Miner* deverá realizar, de

modo a contribuir para a colaboração no desenvolvimento de projectos de engenharia.

O Capítulo IV analisa a implementação do *Miner* e mostra as possíveis aplicações dos seus serviços no contexto da ferramenta criada para gestão de projectos.

Por último, no capítulo V, realizam-se as considerações finais sobre o trabalho proposto e o implementado e a sua contribuição para o projecto CoSKS. Efectuam-se ainda algumas reflexões acerca dos problemas encontrados e apresentam-se as perspectivas futuras.

2. Estado da arte

2.1 Gestão do conhecimento

Foi em 1960 que o filósofo e economista austríaco, Peter Drucker, introduziu o termo “knowledge worker” [1] e a ele viria mais tarde a ser atribuída a paternidade da disciplina de Gestão do Conhecimento. No entanto, só na década de 90 é que se observa um desenvolvimento significativo deste conceito. Autores como Nonaka e Takeuchi [2] estudaram como o conhecimento é produzido, utilizado e difundido no seio das empresas japonesas e de que modo pode contribuir para a difusão de inovação. Vários outros estudiosos, reconhecendo o potencial do conhecimento dentro das empresas, começaram a explorar o melhor modo de gerir o conhecimento, originando assim o aparecimento de vários sistemas computacionais e modelos para gerir o conhecimento produzido.

2.1.1 Definição de gestão do conhecimento

De acordo com a área de trabalho em que o conceito gestão do conhecimento foi aplicado, surgiram diferentes definições. Por exemplo, segundo o empresário Grey a gestão do conhecimento é uma abordagem colaborativa e integrada para a criação, captura, organização, acesso e utilização dos activos intelectuais de uma empresa [3]. Por sua vez, Carl Frappaolo um homem da ciência, propôs que a gestão do conhecimento consiste num aproveitamento da sabedoria colectiva para aumentar a capacidade de resposta e de inovação das empresas [4].

A gestão do conhecimento também foi definida como o conceito segundo o qual a informação é transformada em conhecimento facilmente aplicado pelos indivíduos [5].

Para Kazi, a gestão do conhecimento é a combinação de quatro entidades de conhecimento, pessoas, organizações, informação e tecnologia, de uma forma

sinérgica que permite obter vantagens estratégicas para uma determinada tarefa, num determinado contexto [6].

Uma das definições mais utilizadas para gestão do conhecimento é a elaborada por Nonaka e Takeuchi. Estes autores definiram gestão do conhecimento como um processo para aplicação de uma abordagem sistemática de aquisição, estruturação, gestão e difusão de conhecimento dentro de uma empresa, de modo a permitir práticas de trabalho melhores e mais rápidas, assim como reduzir os custos financeiros [2]. Esta definição é também convergente com a definição utilizada por Davenport e Prusak, sendo que os autores adoptam os mesmos termos e ideias. No entanto, note-se que a definição de Nonaka e Takeuchi prima mais pelo indivíduo, enquanto a de Davenport e Prusak pelo ambiente social [7].

Por ser uma compilação de muitas estratégias, ferramentas e técnicas, é difícil encontrar uma definição simples de gestão do conhecimento. Actualmente considera-se que a gestão do conhecimento engloba tudo o que tem a ver com o conhecimento, deste modo uma boa definição deve incluir todas as perspectivas de criação, representação, armazenamento, distribuição e gestão do conhecimento [8].

2.1.2 Tipos de conhecimento

De acordo com o descrito na secção anterior, verifica-se que não é trivial definir gestão do conhecimento. No que respeita ao conhecimento em si, também muitas teorias surgiram ao longo dos anos, propondo classificações para o conhecimento. Wiig, por exemplo, considerou que existem quatro tipos principais de classificação de conhecimento, factual, conceptual, expectante e metodológico [9].

Mais tarde, Savage na sua análise do conhecimento dividiu-o em seis questões fundamentais: Saber porquê, Saber o quê, Saber como, Saber quem, Saber quando e Saber onde [10].

Na década de 60, Michael Polanyi classificou o conhecimento em apenas dois tipos, tácito e explícito [11], tendo sido seguido por outros autores, nomeadamente Nonaka e Takeuchi [2]. A tabela 2.1 apresenta uma comparação entre os dois tipos de conhecimento supracitados.

Tabela 2.1: Dois tipos de conhecimento [2]

Conhecimento Tácito (Subjectivo)	Conhecimento Explícito (Objectivo)
Conhecimento de experiência (Corpo)	Conhecimento racional (Mente)
Conhecimento simultâneo (Aqui e Agora)	Conhecimento sequencial (Além)

Nonaka considera que o conhecimento tácito é difícil de articular e também difícil de colocar em palavras ou desenhos, sendo por isso altamente pessoal. É difícil de formalizar, tornando-o difícil de comunicar aos outros, estando também muito dependente das acções individuais e do contexto. Quanto ao conhecimento explícito, segundo o mesmo autor, este conhecimento representa o conteúdo que é capturado, como palavras ou imagens, é um conhecimento formal e sistemático. Deste modo, pode ser facilmente comunicado e partilhado, através das especificações de um produto, de uma fórmula científica ou de um programa de computador. Assim, Nonaka deduz que o potencial do conhecimento está na interacção entre o conhecimento tácito e explícito [12].

Juntamente com Takeuchi, Nonaka apresentou um modelo dinâmico de criação de conhecimento que se apoia no facto do conhecimento humano ser criado e expandido através de interacções sociais entre conhecimento tácito e conhecimento explícito [2]. Os autores chamam a esta interacção “conversão de conhecimento”. A figura 2.1 representa os quatro modos de conversão de conhecimento apresentados por Nonaka e Takeuchi. Os autores designaram os quatro modos, no processo de

geração de conhecimento, como Modelo SECI – Socialização, Exteriorização, Combinação e Interiorização, onde:

- **Socialização:** conversão de conhecimento tácito em conhecimento tácito. Partilha de experiências, modelos mentais e habilidades técnicas. A chave para adquirir conhecimento tácito é a experiência.
- **Exteriorização:** conversão de conhecimento tácito em conhecimento explícito. Processo de articulação de conhecimento tácito em conceitos explícitos, em que o conhecimento tácito faz uso de metáforas, analogias, conceitos, hipóteses ou modelos, passando a conhecimento explícito.
- **Combinação:** conversão de conhecimento explícito em conhecimento explícito. Processo de sistematização de conceitos dentro do sistema de conhecimento. Os indivíduos trocam e combinam conhecimento através de diversos meios, tais como, documentos, reuniões, redes de computadores.
- **Interiorização:** conversão de conhecimento explícito em conhecimento tácito. Está relacionada com a prática, para o conhecimento explícito se tornar tácito, é útil que o conhecimento seja verbalizado ou esquematizado em documentos.

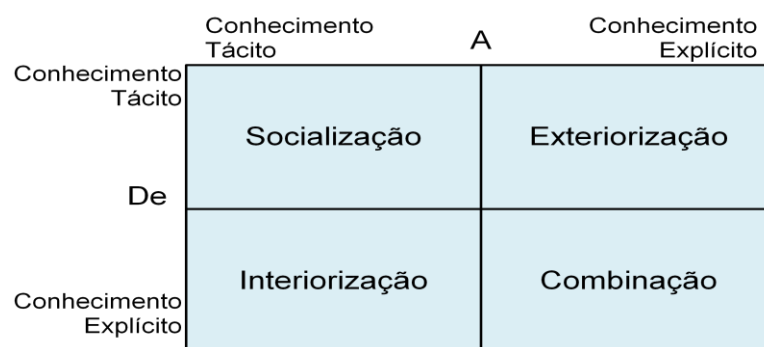


Figura 2.1: Quatro modos de conversão do conhecimento [2]

A interação contínua e dinâmica entre conhecimento tácito e conhecimento explícito vai sendo construída através de mudanças entre os diferentes modos de conversão de conhecimento. Os autores chamaram a este processo a espiral de conhecimento (figura 2.2) e explicam que a criação de conhecimento em empresas é

um processo em espiral, começando no nível individual e movendo-se para cima, por intermédio da expansão de comunidades interactivas, que vão abrangendo outros sectores e departamentos da empresa.

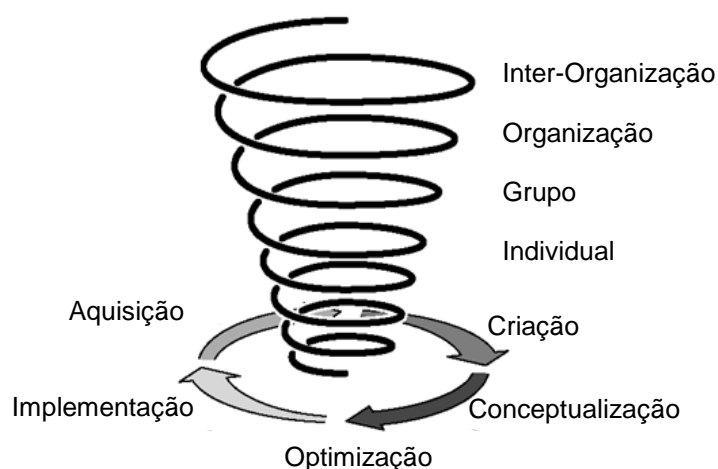


Figura 2.2 - A espiral do conhecimento de Nonaka e Takeuchi [2]

2.1.3 O conhecimento como recurso das empresas

Desde sempre, o capital foi considerado o principal recurso estratégico das empresas. No entanto, com a introdução do conceito de gestão do conhecimento, o conhecimento passou a ser encarado como um recurso de elevado potencial, nomeadamente no sentido de vir a proporcionar vantagens competitivas [13].

Esta visão de organizações baseadas em conhecimento tem sido amplamente explorada e levou à classificação dos recursos de conhecimento disponíveis e que devem ser considerados na implementação de iniciativas de gestão do conhecimento, numa organização. Leonard-Bartion identificou dois tipos de recursos de conhecimento: o conhecimento dos empregados e os sistemas físicos constituídos pelas bases de dados físicas. [14]. Por sua vez, Petrach identificou outros recursos de conhecimento. Para além de reconhecer os empregados como capital humano, adicionou os clientes (capital dos clientes), processos da organização, estruturas da organização e a cultura da organização, referindo-se a estes três últimos como capital da organização [15].

No processo de gestão do conhecimento, os dados das organizações são operados por processadores humanos ou computadores, na realização de actividades de manipulação de dados de modo a criar conhecimento, sob a forma de aprendizagem ou projecções. O processo de gestão do conhecimento é influenciado por uma variedade de factores. Holsapple e Joshi consideram que, para além da criação de conhecimento por processadores de conhecimento, as organizações apresentam outras actividades a operar nos recursos de conhecimento e que também devem ser consideradas, nomeadamente a aquisição, internacionalização, selecção e partilha de conhecimento [16].

Dada a complexidade do processo, vários autores têm vindo a desenvolver modelos e programas que visam a introdução da gestão do conhecimento no seio das organizações.

Skryme admite que um programa de gestão do conhecimento deve ser explícito, sistemático, vital e processual. Portanto, tem de ser um conjunto de actividades com ferramentas e técnicas próprias, que actuam sobre conjunto de dados conhecidos (explícitos e vitais), de uma forma sistemática [17].

Skryme e Amidon (1997) elaboraram uma lista de quatro estratégias distintas que devem orientar a implementação de uma filosofia de gestão do conhecimento, de acordo com o tipo de organização. A primeira estratégia dita que o conhecimento deve ser gerido como um activo da organização (nuclear) e, deste modo deve ser apoiado e protegido, sendo aplicado para organizações com elevada capacidade intelectual (e.g. indústria farmacêutica) A segunda estratégia relaciona-se com o facto de ser necessário realçar a capacidade dos produtos e serviços, ou seja, o conhecimento deve torná-los mais atractivos e é apropriada para empresas de alta tecnologia. Por outro lado, a terceira estratégia, mais comum em empresas de consultoria, postula que o conhecimento deve ser gerido como um negócio.

Finalmente, na quarta estratégia, direccionada principalmente a serviços financeiros e indústrias de manufactura em geral, a gestão do conhecimento deve permitir realçar os processos de negócios e gestão, identificando o conhecimento e tornando-o acessível sempre que necessário [18].

Elliot e O'Dell, em 1999, sugeriram que os quatro elementos chave, imprescindíveis para o sucesso da gestão do conhecimento numa empresa, são a cultura, a tecnologia, as infra-estruturas e avaliação do desempenho. Actualmente, estes elementos são admitidos como peças fundamentais na gestão do conhecimento. A cultura influencia todos os comportamentos dos indivíduos e a tecnologia promove as ferramentas que possibilitam o desenvolvimento das comunicações. Por sua vez, as infra-estruturas constituem as regras de suporte aos processos da empresa. Por último, a avaliação, é fundamental porque permite avaliar os desempenhos e decidir a necessidade de corrigir estratégias [19].

2.2 Gestão do conhecimento colaborativa

Da discussão apresentada na secção anterior, resulta a ideia de que a gestão do conhecimento depende, inevitavelmente, das capacidades colectivas dos indivíduos que partilham o conhecimento, durante o processo. Actualmente, a colaboração tem vindo a ser definida, por vários autores, como um dos três pilares da gestão do conhecimento, a par das etapas de consulta e partilha de dados [20].

2.2.1 O conceito de mercados de conhecimento

Em 1998, os autores Davenport e Prusak, reconhecendo a importância de estudar a etapa de colaboração, nos processos de gestão do conhecimento, introduziram o conceito de “mercados de conhecimento” [7]. Definiram estes mercados como *locais virtuais, onde vários indivíduos devem interagir para trocar conhecimento* e efectuaram um estudo detalhado de como os mercados devem funcionar para serem

eficientes. Assim, identificaram como participantes no mercado os *vendedores*, os *compradores* e os *correctores*. Os compradores foram definidos como indivíduos que pretendem resolver uma determinada situação complexa, para a qual não há uma solução simples, enquanto os vendedores são indivíduos na posse de conhecimento substancial sobre um determinado processo ou área, passível de ser partilhado. Por último, os correctores são os moderadores do processo de partilha de conhecimento, responsáveis por realizar contactos entre compradores e vendedores.

Outro aspecto dos mercados do conhecimento estudado pelos autores, é a questão do sistema de preços a praticar, ou seja, como as empresas podem trocar o seu capital intelectual. Davenport e Prusak consideram que para o mercado operar eficientemente é necessário que factores como a reciprocidade, o respeito, o altruísmo e a confiança, sejam apresentados pelos intervenientes do mercado.

Para facilitar esta questão do pagamento, Simard propôs um modelo de mercado com uma natureza sequencial, cíclico e composto por nove estágios (figura 2.3).

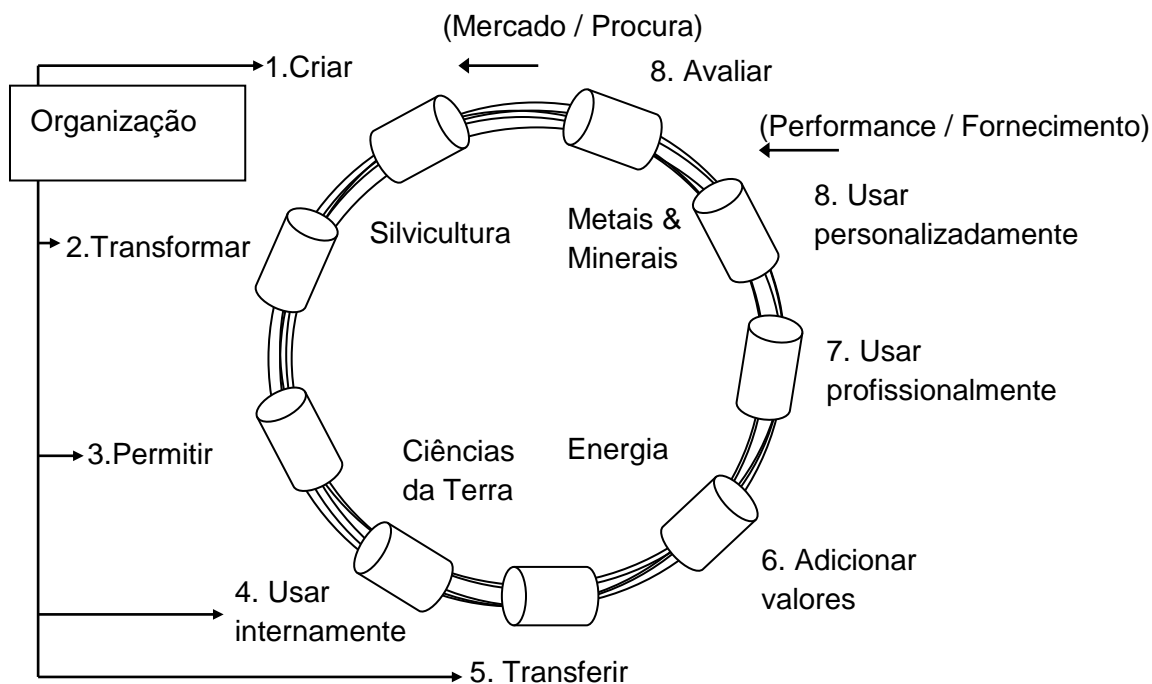


Figura 2.3 - Modelo para mercado do conhecimento proposto por Simard [21]

Os primeiros cinco estágios (produção e transferência de conhecimento) ocorrem no interior da empresa e os últimos quatro estágios são externos à empresa (intermediários e clientes). Neste tipo de mercado cíclico, ambos os processos de pedido e fornecimento de informação estão contemplados para cada membro do mercado (compradores e vendedores), garantindo assim que todos os membros disponibilizem e recebam conhecimento [21].

2.2.2 Tecnologias que suportam a colaboração em projectos de gestão do conhecimento

A necessidade das empresas actuarem em locais geograficamente dispersos e de poder utilizar todos os participantes, levou à procura de meios eficientes para facilitar o contacto a longa distância. Este tem sido um grande desafio para as empresas e neste sentido, a busca de novas ferramentas e sistemas tecnológicos que facilitem o trabalho colaborativo tem sido vital [22].

Para facilitar o processo colaborativo, a gestão do conhecimento suporta a ideia de um ambiente de trabalho de elevado desempenho, um modo de trabalho que permita estimular os empregados, parceiros ou clientes de uma empresa a criar, obter e organizar o conhecimento e tornarem-se mais colaborativos e inovadores [12]. A implementação destes ambientes de trabalho tem sido facilitada pelo aumento da velocidade das redes e as inovações dos computadores actuais também têm facilitado a utilização de ferramentas e sistemas de capacidade elevada.

As ferramentas usadas em processos de gestão do conhecimento são ferramentas que automatizam e suportam a manipulação das bases de dados disponíveis [23]. Segundo Tyndale, estas ferramentas devem permitir que os processos de conhecimento “tomem decisões” [24]. Alguns exemplos de ferramentas utilizadas em gestão de dados são intranets, sistemas de gestão de conteúdo, bases de dados,

sistemas de armazenamento de dados, sistemas de DM, videoconferência [25]. Muitas destas tecnologias são utilizadas tanto em gestão do conhecimento individual como em trabalhos colaborativos. Por outro lado, existem outras que constituem plataformas exclusivamente de colaboração e comunicação entre profissionais, como é o caso do *Groupware* (aplicações de trabalho em grupo ou entre grupos) [24].

Os autores Bafoutsu e Mentzas propuseram uma classificação das ferramentas de gestão do conhecimento e a sua relação com o nível de funcionalidade para colaboração [26]. As quatro categorias, de acordo com os autores, estão distribuídas num gráfico a duas dimensões (figura 2.4), com o eixo dos X a representar os níveis de colaboração, começando com os serviços que promovem a simples comunicação e terminando nos serviços que promovem a colaboração e interacção humana. Por outro lado, o eixo dos Y representa a capacidade de gestão de documentos, começando na simples partilha de documentos e terminando no trabalho síncrono num documento.

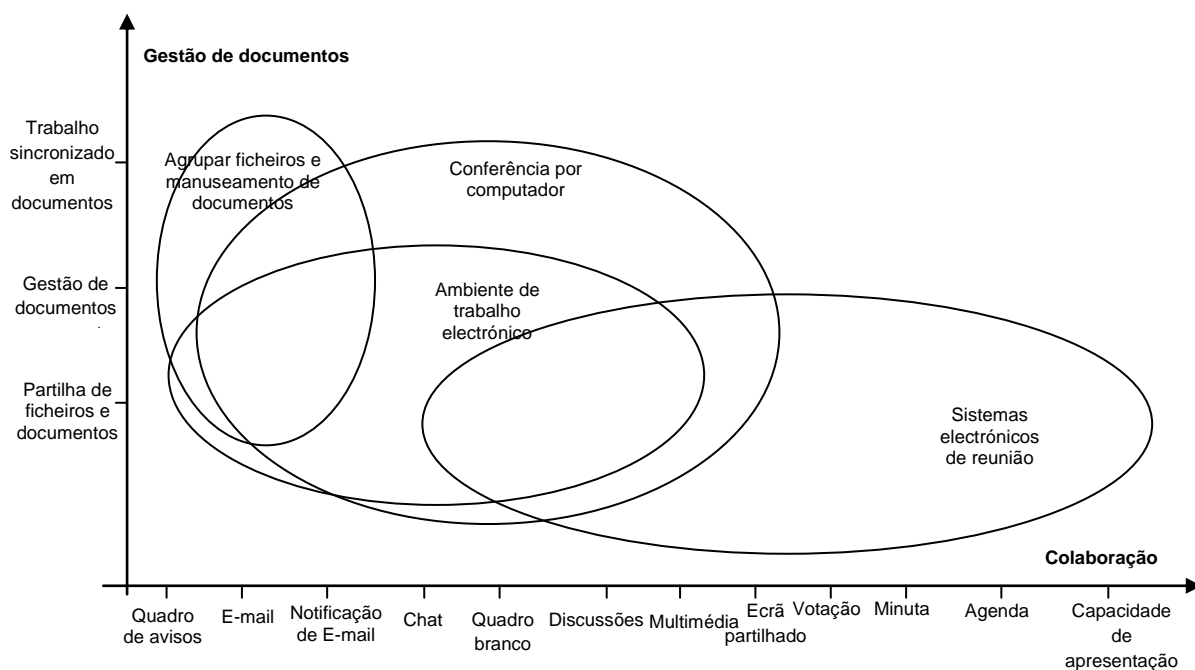


Figura 2.4 - Categorização de ferramentas para conhecimento [26]

Ainda neste estudo, os autores elaboraram uma lista de serviços colaborativos constituída por 47 sistemas, alguns deles disponíveis comercialmente, outros permanecendo como protótipos. Estes sistemas disponibilizam vários serviços colaborativos tais como *bulletin board*, e-mails, chat e videoconferência.

2.2.3 Ineficiências da gestão do conhecimento colaborativa

São cada vez mais frequentes os projectos sobre gestão do conhecimento descritos na literatura, sendo que nem todos têm sido bem sucedidos. De acordo com Davenport e Prusak, há vários factores que podem estar na origem destes projectos falhados. O facto de existir informação incompleta sobre um determinado assunto ou de não ser possível encontrar a informação pretendida, a assimetria de conhecimento, nomeadamente por uma das partes ter mais informação a partilhar do que outra, têm constituído entraves para o processo de colaboração.

Há ainda aspectos culturais das empresas, pré-estabelecidos de longa data e que podem constituir barreiras à gestão do conhecimento colaborativa. Long e Fahey [27] identificaram quatro modos da cultura de uma empresa influenciar os comportamentos centrais na criação, partilha e utilização de conhecimento. O primeiro destes aspectos contempla o facto de as empresas possuírem uma pré-definição do tipo de conhecimento que é importante o que leva à tendência para excluir qualquer novo tipo de conhecimento. Outro aspecto relaciona-se com o facto da cultura de empresa definir as relações entre os indivíduos e atribuir aos indivíduos o tipo de conhecimento que lhe diz respeito. O terceiro modo prende-se com o facto da cultura influenciar o contexto para a interacção social, determinando as regras e as práticas (e.g. o formato e frequência das reuniões ou o uso apropriado de e-mails). Estas regras culturais moldam o modo como as pessoas interagem e acabam por ter um impacto substancial na partilha do conhecimento. Por último, a cultura é também responsável pelo modo como o novo conhecimento vai ser

processado. Segundo os autores, a identificação destes factores culturais como barreiras para o processo de gestão do conhecimento, é o primeiro passo para o desenvolvimento de uma estratégia que permita uma gestão do conhecimento muito mais eficaz.

Segundo os mesmos autores, a identificação destes factores é um grande passo para a resolução dos problemas encontrados nos projectos colaborativos e pode contribuir para o desenvolvimento de projectos colaborativos de gestão do conhecimento mais eficientes.

2.3 DM como ferramenta no processo de gestão do conhecimento

O processo de gestão do conhecimento é interactivo, envolvendo várias ferramentas que permitem adquirir conhecimento. Uma dessas ferramentas é o DM, que permite analisar grandes quantidades de dados e identificar padrões válidos nos dados originais.

2.3.1 Etapas do processo de descoberta de conhecimento

Em 1996, Fayyad e colaboradores identificaram várias etapas no processo de descoberta de conhecimento (figura 2.5) [28]:

1. Reconhecimento do domínio da aplicação e o conhecimento que será relevante, ou seja, identificação do objectivo do processo.
2. Selecção de um conjunto de dados alvo, nos quais o processo de gestão do conhecimento será aplicado.
3. Limpeza e processamento dos dados, sendo que nesta etapa, as operações incluem por exemplo a remoção de ruídos, a escolha de estratégias para lidar com campos em falha ou contabilizar as alterações que possam ser realizadas.
4. Redução e projecção de dados, utilizando técnicas que representem os dados em função do objectivo em questão.

5. Adequação dos objectivos em questão a um método de DM (sumarização, classificação, regressão, *clustering*).
6. Realização da análise exploratória, escolhendo um algoritmo de DM, assim como os parâmetros que serão adaptados.
7. Aplicação do DM, em busca de padrões de interesse em representações particulares, tais como diagramas em árvore, regressões ou *clustering*.
8. Interpretação dos padrões obtidos e eventualmente repetir o processo até esta etapa.
9. Actuação sobre o conhecimento descoberto, incorporação noutra sistema para posterior acção ou simplesmente documentação.

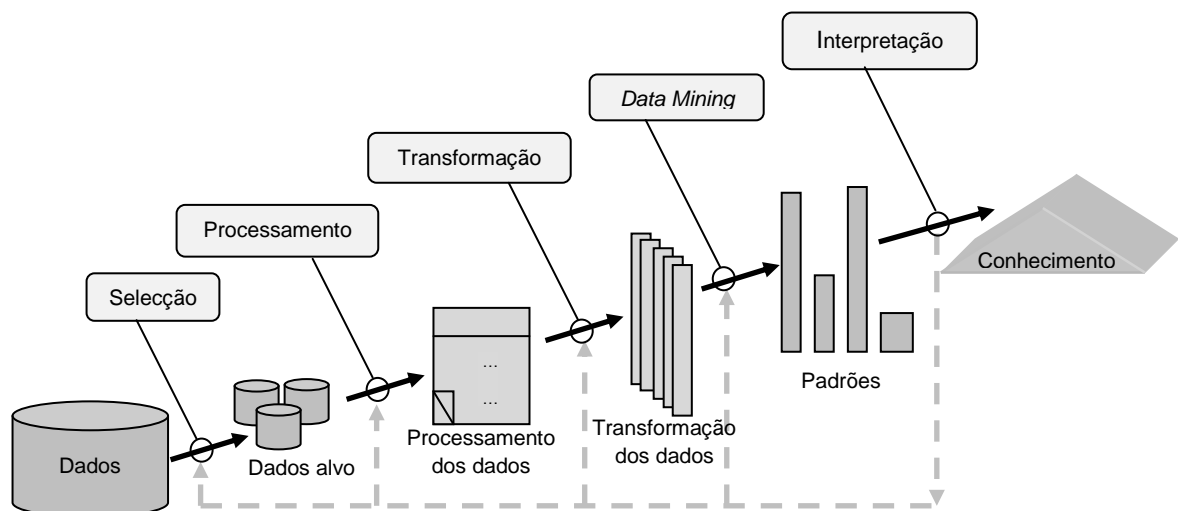


Figura 2.5 - *Data mining* e o processo *knowledge discovery in databases* [28]

Não se deve confundir DM e *knowledge discovery in databases* (KDD). Enquanto este último trata-se do processo de encontrar informações úteis e padrões nos dados, o DM utiliza algoritmos para extrair informação e padrões determinados pelo processo de KDD.

2.3.2 Tarefas de DM

À medida que o conceito de gestão do conhecimento ganhou importância, a criação de ferramentas de DM sofisticadas para a realização das análises de dados

tornou-se cada vez mais uma necessidade. Actualmente, com o desenvolvimento destas ferramentas, é possível realizar algumas tarefas, a partir dos dados existentes, das quais se destacam [29, 30]:

1. Processamento de dados: de acordo com os objectivos pretendidos no processo KDD, a análise dos dados pode incluir a selecção, filtração, agregação, amostragem, limpeza ou transformação dos dados.
2. Predição: dado o valor de um determinado dado e um modelo de predição, é previsto o valor de um atributo específico.
3. Regressão: na presença de um conjunto de dados, é efectuada a análise da dependência de alguns dos valores em função de outros.
4. Regras de classificação: dado um conjunto de classes predefinidas, é determinado a qual dessas classes um conjunto de dados específico pertence.
5. Regras de *clustering*: dado um conjunto de dados, estes são agrupados de acordo com o seu grau de semelhança, permitindo uma melhor identificação dos mesmos.
6. Regras de associação (sumarização): dado um conjunto de dados, são identificadas as relações e padrões existentes entre atributos.
7. Visualização dos dados: a visualização desempenha um papel importante na descoberta de conhecimento compreensível e interpretável pelos seres humanos.
8. Análise exploratória de dados: é a exploração interactiva de um conjunto de dados que não revelam relações evidentes, de modo a identificar padrões interessantes.
9. *Similar Time Series*: consiste na procura de eventos semelhantes num determinado período de tempo e determinação de características que levam à ocorrência desse acontecimento.

2.3.3 Técnicas de DM

De modo a possibilitar a execução do conjunto de tarefas descritas anteriormente, têm sido desenvolvidas técnicas de DM. Cada técnica poderá realizar uma ou mais tarefas. Contudo, em qualquer uma das técnicas, a existência de algoritmos tem como principal objectivo extrair informação e encontrar padrões semelhantes nos dados.

Inicialmente, as técnicas de DM baseavam-se em métodos estatísticos, tais como regressões não paramétricas e ferramentas como validação-cruzada. No entanto, em paralelo com o crescimento dos processos de DM, a ciência computacional experimentou um avanço nos modelos de redes neuronais e novos algoritmos, tais como o *clustering*, que permitiram executar rapidamente as análises estatísticas tradicionais em conjuntos grandes de dados [31]. De entre as muitas técnicas existentes que realizam DM, destacam-se as seguintes [29]:

- *Case Based Reasoning* (CBR): a técnica CBR tenta resolver novos casos (novos problemas) baseando-se em soluções apresentadas na resolução de casos passados. Quanto maior for o número de casos anteriormente resolvidos, e com maior número soluções eficazes, maior será a probabilidade de encontrar uma solução para o novo caso em estudo. Também, à medida que novos casos vão sendo resolvidos, maior será a base de dados possível para uso futuro. Resumindo, CBR é uma técnica de resolução de problemas que utiliza a experiência para resolver novos problemas. O CBR pode ser dividido em quatro fases: (i) recuperação (após apresentado um novo problema ao sistema, é realizada uma recuperação na base de casos de modo a apresentar-se o caso mais parecido com o caso em questão); (ii) reutilização (a partir do caso recuperado é feita a reutilização da solução associado a esse caso e, aplicada ao novo caso); (iii) revisão (quando a solução do caso

semelhante não pode ser aplicado como sendo solução para o novo caso, há necessidade de verificar e aplicar apenas as características da solução que sejam úteis para resolução do novo caso); e (iv) retenção (armazenar o novo caso e respectiva solução para futuras recuperações).

- **Redes Neurais:** Trata-se de uma técnica que utiliza uma abordagem computacional com o desenvolvimento de estruturas matemáticas com capacidade de *aprendizagem*. Uma rede neuronal consiste num número de elementos interligados (neurónios) que possuem entrada, saída e processamento. De modo a construir um modelo com redes neurais é necessário, inicialmente realizar um treino da rede recorrendo a um conjunto de dados e, nesse caso é utilizada esta rede para realizar predições. Existem alguns problemas quando se pretendem usar redes neurais, nomeadamente pelo facto de serem difíceis de compreender e só aceitarem dados numéricos, pelo que todos os dados têm de ser convertidos em dados numéricos.
- **Árvores de decisão:** uma árvore de decisão utiliza uma estratégia de “dividir para conquistar”. Tenta decompor um problema complexo em problemas mais simples e, isso é feito recursivamente até se chegar a uma resposta/solução. As árvores de decisão são constituídas por nós (que representam os atributos dos dados) e por ramos (que representam os possíveis caminhos ao longo da árvore). Em cada nó da árvore é colocada uma questão e, dependendo da resposta, segue-se um ramo da árvore até se chegar a uma resposta pretendida.
- **Regras de Indução:** esta técnica refere-se à detecção de tendências que se verificam dentro de grupos de dados.
- **Algoritmos Genéticos:** algoritmos genéticos representam estratégias para otimizar algoritmos baseados nos princípios observados na evolução natural

das espécies. De um conjunto de soluções existentes para um determinado problema, que competem umas com as outras, as melhores soluções são seleccionadas e combinadas entre si. O objectivo passa por encontrar a melhor solução para o problema, de um modo semelhante ao processo de evolução de uma população de organismos. Os algoritmos genéticos são usados no processo DM para formular hipóteses sobre dependência entre variáveis, na forma de regras de associação

- Conjuntos *Fuzzy*: os conjuntos *Fuzzy* são a chave para a representação e processamento de incertezas (imprecisão, inespecificidade, inconsistência, etc) existentes nas bases de dados. Por outro lado, a lógica *Fuzzy* constitui uma ferramenta poderosa para lidar não só com incertezas, podendo também ser útil no desenvolvimento de modelos de incerteza dos dados.

2.3.4 Aplicações do DM

Actualmente encontram-se descritos vários exemplos concretos da sua utilização nas mais diversas áreas. Nomeadamente, ao nível do *marketing*, o DM pode ser utilizado para identificar diferentes grupos de clientes e prever o seu comportamento. Por exemplo, em 1996, Agrawal aplicou um sistema de modo encontrar padrões de associação na venda de produtos num supermercado, ou seja, tentou avaliar se a compra de determinado produto implica a compra de outro produto [32]. Também ao nível da indústria, o DM tem permitido encontrar respostas importantes. Um destes exemplos é o sistema de solução de problemas Cassiopée, desenvolvido, pela General Electric e a SNECMA, o qual foi aplicado por três grandes companhias aéreas europeias para diagnosticar e prever problemas para o Boeing 737.

Actualmente, também na área da medicina o DM tem sido aplicado. Em 2006, Freitas aplicou DM (classificação, regressão e *clustering*) em dados hospitalares. Foram aplicadas estratégias de classificação dependentes de vários custos,

económicos ou não, associados principalmente a testes médicos. Neste estudo, obtiveram-se “novos conhecimentos” tais como, a diminuição de internamentos de longa duração, o aumento do número de diagnósticos secundários. Segundo o autor, os conhecimentos obtidos podem ser indicadores importantes para a análise de gestão e de planeamento hospitalar [33].

Por outro lado, Luan, na área da educação, propôs a utilização de técnicas de *clustering* para a análise compreensiva de características de estudantes e de técnicas de predição para estimativa de probabilidades de comportamentos, tais como a persistência, retenção e sucesso escolar [34]. Um exemplo de aplicação proposto, neste projecto, inclui a utilização de DM para prever a possibilidade de cada estudante matriculado num determinado colégio (College of Silicon Valley) voltar a estudar. A lista obtida, com os nomes de estudantes com probabilidade reduzida de voltar a estudar, é utilizada para permitir gerir/intervir na situação destes alunos. Neste artigo também se discute a aplicação de DM em educação superior.

2.3.5 Text Mining

O *text mining* (TM), à semelhança do processo de DM, identifica relações entre conjuntos de dados. No caso do TM o que se pretende é a descoberta de informação nova, pela extracção automática de informação de várias fontes de texto escrito [35].

A aplicação de TM permite realizar diversas acções sobre um documento, tais como, identificar o idioma do texto, extrair e seleccionar características do texto (datas, palavras chave, nomes), sumarização do documento. Sempre que se pretende retirar informação de um determinado documento (i.e. aplicar TM) é necessário ter em conta alguns procedimentos básicos no processamento dos textos [36], nomeadamente no que respeita a:

1. Propriedades das palavras: deve-se estudar a relação entre a forma das palavras e o seu significado. Por exemplo, a palavra “banco” pode representar uma instituição financeira, ou um assento.
2. Remoção de *stop words*: palavras que não contém informação relevante na aplicação do TM, por exemplo na língua portuguesa temos: *a, à, de, e, ou, com, etc..*
3. *Stemming*: processo de redução de todas as palavras a um termo mínimo, possível de comparação. Por exemplo a palavra “informação” poderá ser reduzida a “info”, deste modo, quando se analisa o texto, as palavras informar, informador, também serão reconhecidas.
4. Frequência de palavras: determinação da frequência de uma palavra ou uma sequência de palavras existentes no texto.

Os procedimentos descritos anteriormente fazem parte da etapa de preparação do processo de TM. Segundo Liddy, o processo de TM é constituído por três fases: preparação (limpeza do dados), processamento (utilização de um algoritmo de TM para processar os dados preparados) e análise (avaliação do conhecimento descoberto e avaliação da sua importância) [37].

De acordo com Cios e colaboradores, se os processos de TM forem bem implementados, podem constituir uma mais-valia principalmente no que respeita a recuperação de informação [38].

2.3.6 Análise de ferramentas de DM

Os primeiros softwares para realização de DM começaram a ser desenvolvidos em meados da década de 90 em ambientes académicos. Face ao rápido desenvolvimento da área, actualmente encontram-se disponíveis várias ferramentas, quer académicas quer comerciais.

Dada a impossibilidade de analisar todas as ferramentas de DM actualmente disponíveis, optou-se por analisar cinco destas ferramentas, todas de licença pública e que se encontram entre as mais populares devido às suas qualidades:

- **MiningMart:** ferramenta gráfica para pré-processamento de dados e DM em bases de dados desenvolvida com o objectivo de realizar o pré-processamento de grandes volumes de dados [39].
- **Orange:** ferramenta que inclui uma grande variedade de padrões e algoritmos de DM, além de rotinas para entrada e manipulação de dados [40].
- **Tanagra:** ferramenta de DM com propósitos académicos e de investigação que permite a utilização de várias técnicas de DM tais como a análise exploratória e visualização de dados [41].
- **Weka:** ferramenta de *software* mais utilizada para efectuar tarefas de DM. Os seus algoritmos podem ser aplicados directamente a um conjunto de dados a partir da sua própria interface ou utilizando o próprio código Java. Contém ferramentas de pré-processamento, classificação, regressão, *clustering*, regras de associação e visualização [42].
- **RapidMiner:** consiste num ambiente de KDD que fornece uma rica variedade de métodos. O RapidMiner é reconhecido como uma versão melhorada do Weka, incluindo todas as suas funcionalidades [43].

Vários autores têm estudado os factores de decisão que devem ser considerados para seleccionar as ferramentas mais adequadas para determinado projecto de *software* [44, 45].

Na tabela 4.2 caracterizam-se as cinco ferramentas introduzidas anteriormente de acordo com os critérios de funcionalidade importantes para o trabalho em estudo. Os critérios escolhidos incluíram a **linguagem** utilizada no desenvolvimento do software, a **conectividade a bases de dados**, em particular a MySQL, a **interface gráfica**, a

capacidade de realizar **TM**, assim como as **tarefas de DM** permitidas. Na interface gráfica consideraram-se três características principais, a **representação gráfica** uma vez que esta permite uma melhor interpretação assim como um aparência mais *user-friendly* dos dados, a **visualização de dados** que avalia a capacidade de representar dados por intermédio de gráficos, tabelas ou outras estruturas semelhantes e a **gestão de dados**, nomeadamente a capacidade da ferramenta disponibilizar opções que permitem a realização de operações de manipulação de dados tais como remover ou alterar linhas, colunas, etc..

Tabela 2.2: Ferramentas DM (adaptado de [46])

Ferramenta	Linguagem	Conectividade a base de dados	TM	Interface Gráfica			Tarefas DM			
				Representação gráfica	Visualização de dados	Gestão de dados	Classificação	Regressão	Clustering	Regras de associação
MiningMart	Java	S		S	B	A	B	B	N	N
Orange	C++	N	✓	S	A	A	M	N	M	M
Tanagra	C++	N		N	A	A	A	M	A	A
Weka	Java	S	✓	S	A	A	A	A	A	A
RapidMiner	Java	S	✓	S	A	A	A	A	A	A

Legenda: N - não suporta B - suporte básico M - suporte médio A - suporte avançado

2.4 Síntese

Ao longo deste capítulo foram abordados diversos conceitos importantes relacionados com a gestão do conhecimento. Justificou-se que a colaboração e a

utilização de DM para explorar o conteúdo dos dados armazenados podem aumentar o desempenho de empresas e instituições em várias áreas.

No presente trabalho, os conceitos colaboração e DM são explorados no sentido de obter novas informações a partir de dados armazenados, num ambiente colaborativo orientado para projectos. Assim, recorrem-se a algumas tarefas de DM nomeadamente a predição, a classificação e a associação, utilizando principalmente a técnica de árvores de decisão. A análise das ferramentas de DM disponíveis levou à escolha do RapidMiner, por permitir a ligação com base de dados geridas pelo MySQL e ser desenvolvida na linguagem de programação Java, o que possibilita uma futura integração com a interface CoSKS. Por outro lado, engloba um vasto conjunto de tarefas de DM e TM e possui uma interface gráfica bastante acessível, quando comparada com as outras ferramentas. Outro aspecto positivo é o facto de integrar as bibliotecas do Weka, aumentando assim o seu conjunto de algoritmos e tarefas de DM.

3. Requisitos e modelo conceptual

Este capítulo apresenta os requisitos que conduziram o desenvolvimento do *Miner* e o respectivo modelo conceptual que suportou a sua implementação. Numa primeira fase são discutidos os requisitos mínimos e necessários para que o *Miner* funcione em harmonia com a arquitectura CoSKS. Em seguida relacionam-se três conceitos fundamentais e que de algum modo contribuem para o desenvolvimento de projectos de engenharia: colaboração, conhecimento e raciocínio e discute-se o modo como cada um dos elementos é importante para o *Miner*. Por último descreve-se o conceito *Miner*, a interacção do *Miner* dentro do CoSKS, os respectivos serviços do *Miner* e sua contribuição no desenvolvimento dos projectos.

3.1 Termos

Os termos mais utilizados neste trabalho encontram-se definidos na tabela 3.1.

Tabela 3.1: Termos básicos e definições

Termo	Definição
Projecto	Conjunto de questões, tarefas, reuniões realizado por um conjunto de pessoas/empresas, de um modo organizado e colaborativo, num determinado período de tempo, com o objectivo de produzir/realizar/atingir alguma coisa.
Questão	Qualquer desvio, problema ou assunto que possa surgir ao longo do tempo de vida de um projecto. Cada questão será analisada e discutida nas reuniões de projecto. Para cada questão uma ou mais tarefas são realizadas de modo a resolver a mesma.
Tarefa	Acto que permite resolver um ou mais problemas (questões) dos projectos. Cada tarefa terá uma pessoa como responsável principal e um conjunto de pessoas capazes de efectuar a tarefa.
Actor	Pessoa participante no projecto.
Agenda	Lista de tópicos a discutir numa determinada reunião. Cada tópico relaciona-se com uma questão específica.

Reunião	Acto colaborativo que ocorre durante o período de vida do projecto. As reuniões são efectuadas com o intuito de discutir as fases do projecto, problemas que surgem ao longo do projecto e encontrar as respectivas soluções. Qualquer actor do projecto pode participar nas reuniões. É necessário que exista pelo menos um elemento, em cada reunião, capaz de elaborar um relatório sobre os assuntos discutidos na mesma (documento).
Minuta	Rascunho inicial descrevendo os assuntos discutidos durante uma reunião.
Elemento de conhecimento	<i>Peça</i> de conhecimento focada em tópicos específicos. Inclui documentos sobre experiências adquiridas, melhores práticas, manuais. No contexto do CoSpaces, os projectos, agendas, reuniões, tarefas, questões são também considerados elementos de conhecimento.
Ponto de decisão	Ponto de controlo em determinado momento do projecto, no qual se tentam encontrar soluções para os problemas que possam estar a ocorrer nesse mesmo momento.
Documento	Relatório de tarefas, reuniões ou qualquer ficheiro que contribua para o projecto.
Dados	Conjunto de registos que podem ser processados em informação (e.g. atributos de base de dados).
Informação	Conjunto de dados processados que adquirem significado num determinado contexto.
Conhecimento	Conjunto de informações que, aliadas a experiências, facilitam o processo de tomada de decisão.
Miner	Componente responsável pelo processo de DM, também conhecido como extracção de informação de bases de dados, sendo um processo complexo de identificação de padrões válidos, existentes nos dados. Este tenta contribuir de algum modo para o processo de colaboração ao longo do desenvolvimento de um projecto.
Companion	Componente proactivo, reactivo e autónomo que permite monitorar, alertar e apoiar os utilizadores no processo de tomada de decisão.

3.2 Os requisitos do projecto CoSpaces

O presente trabalho é desenvolvido no contexto do projecto CoSpaces, especificamente no componente de suporte ao conhecimento, denominado CoSKS. O CoSKS faz parte da *CoSpaces Software Framework* (CSF), um ambiente de engenharia colaborativo capaz de suportar a colaboração em tempo real entre equipas dispersas geograficamente. O CSF faz uso de tecnologias avançadas, tais como realidade virtual, tecnologias móveis e serviços Web de modo a criar espaços colaborativos de apoio à decisão humana. As funcionalidades ligadas ao conhecimento devem estar habilitadas para as ontologias e para a realização de técnicas de DM, de modo a permitir o apoio ao desenvolvimento dos projectos.

A abordagem utilizada pelo CoSKS centra-se em dois conceitos: questões/problemas e decisões. Com esta abordagem pretende-se acompanhar todas as questões que surgem durante o desenvolvimento de um projecto, assim como as respectivas decisões que forem tomadas a fim de resolvê-las. O objectivo é fornecer às equipas do projecto uma melhor compreensão dos problemas comuns, que ocorrem com frequência, em projectos e também indicar estratégias (tarefas e acções) para superá-los. Deste modo, é possível melhorar o desempenho diário das equipas do projecto, fornecendo casos históricos de projectos anteriores e informação sobre o modo como estes casos foram resolvidos, reduzindo assim o tempo dispendido na procurando de boas opções em situações específicas.

O processo de pesquisa de informação pode ser realizado na fase de iniciação de um projecto (e.g. o gestor do projecto pode adquirir conhecimentos relevantes sobre projectos bem sucedidos na mesma área, tais como desvios, principais problemas e soluções, etc.) ou durante as reuniões (e.g. os participantes podem ter conhecimento mais preciso sobre o seu projecto, conhecendo os riscos associados a uma determinada decisão). Por outro lado, o CoSKS também precisa manter o

controlo de decisões importantes tomadas durante as reuniões. Este controlo é realizado pela elaboração de agendas que incluem os temas a serem discutidos e de minutas onde são registadas as questões levantadas assim como as soluções propostas e decisões tomadas durante a reunião.

A figura 3.1 mostra o CoSKS inserido na arquitectura CSF, evidenciando os componentes mais importantes para o CoSKS, o BSCW (*Basic Support for Cooperative Work*) e o portal. O BSCW é uma tecnologia que gere e armazena ficheiros e elementos de conhecimento. O portal permite aos utilizadores aceder às funcionalidades disponibilizadas pelo CoSKS.

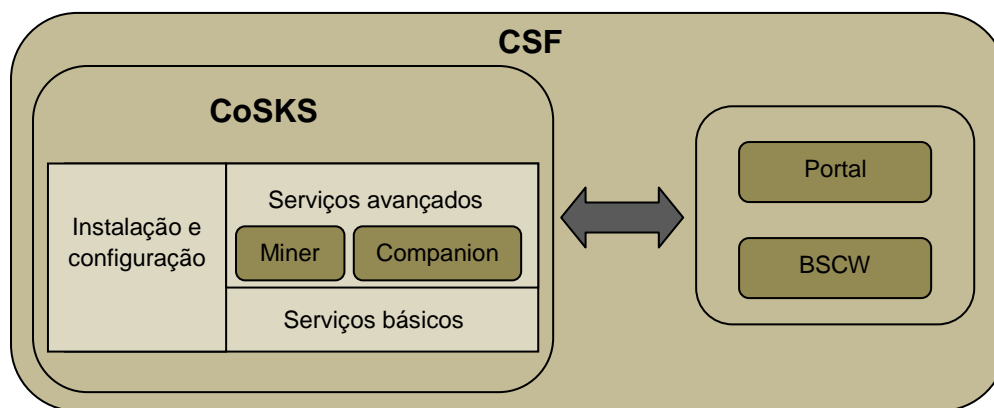


Figura 3.1: Visão geral da arquitectura do CSF

Como se observa na figura 3.1, o CoSKS é constituído por três módulos: instalação e configuração, serviços básicos e serviços avançados.

O módulo de instalação e configuração manipula as necessidades relacionadas com a instalação e configuração do componente CoSKS (por exemplo, criação de utilizadores/senhas/endereços de servidores, etc.).

O módulo de serviços básicos constitui o suporte para o funcionamento normal do CoSKS. Neste módulo estão incluídos, por exemplo, serviços de indexação de elementos de conhecimento, pesquisa na base ontológica, armazenamento/recuperação de objectos de conhecimento, etc..

Por último, o módulo de serviços avançados do CoSKS, do qual fazem parte o *Miner* e *Companion*, tem como objectivo alargar e enriquecer as capacidades do CoSKS. De facto, os serviços destes dois módulos tem como objectivos gerir, apoiar, captar e capitalizar conhecimento proveniente de uma série de reuniões colaborativas.

Neste trabalho foi desenvolvido o módulo *Miner*, que integra os serviços avançados do CoSKS. Neste sentido, será necessário, inicialmente, analisar o funcionamento do CoSKS, assim como as suas principais necessidades, de modo a enquadrar a implementação *Miner*.

Os principais requisitos que devem ser considerados para o CoSKS podem ser divididos em três categorias: requisitos funcionais, arquitecturais e técnicos (figura 3.2).

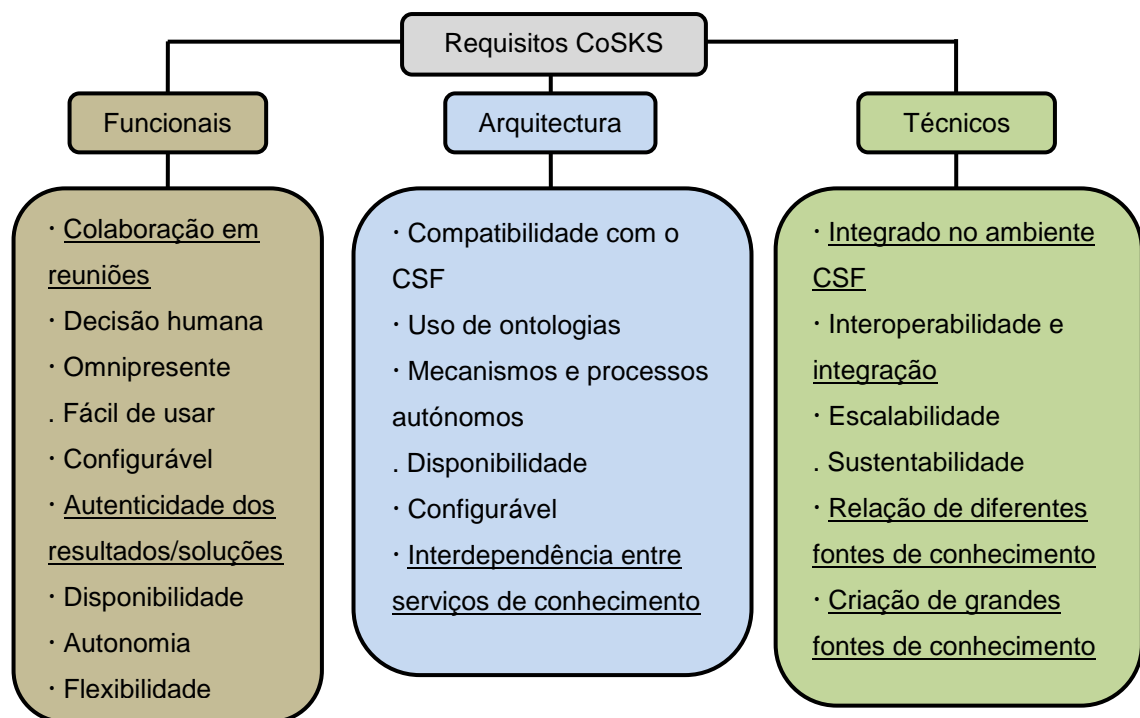


Figura 3.2: Requisitos do CoSKS (adaptado de [47])

Todos os requisitos apresentados são importantes para o desenvolvimento e bom funcionamento do CoSKS. Contudo, no que respeita ao *Miner*, alguns requisitos adquirem maior importância. O *Miner* é um módulo que necessita de grande quantidade de dados para produzir informação relevante. Assim, qualquer requisito

que contribua de algum modo para gerar, captar, gerir ou capitalizar dados será inevitavelmente um requisito do *Miner*. Neste sentido, alguns requisitos mínimos necessários incluem a **colaboração**, a **integração** assim como todos os que **envolvam a obtenção de dados** que possa originar informação, tais como, a **interdependência entre os vários serviços de conhecimento**, a **relação entre diferentes fontes de conhecimento** e a **aquisição e gestão de grandes fontes de conhecimento**.

No que respeita à **colaboração**, é importante que, numa série de reuniões colaborativas, o CoSKS seja capaz de registar, classificar e integrar os dados resultantes dessas reuniões. Por sua vez, os dados adquiridos ao longo das reuniões colaborativas poderão ser utilizados pelo *Miner* para criar informação, de modo a responder a eventuais problemas que possam surgir em futuros projectos, ou simplesmente para encontrar padrões nos diferentes projectos existentes.

Quanto à **integração**, pretende-se que o CoSKS comunique com outros ambientes de tecnologias de informação (por exemplo, a integração com o BSCW), de modo a adquirir mais dados. Por outro lado, as empresas que venham a utilizar o CoSKS, eventualmente terão em funcionamento os seus próprios sistemas computacionais, pretendendo continuar a utilizá-los para o armazenamento dos seus dados, no CoSKS. Deste modo, é necessário existir um mecanismo disponível que seja adequado, fácil de usar e flexível, de modo a apoiar a integração de tais sistemas.

Também é necessário que o CoSKS consiga gerir de forma adequada os dados obtidos. Nomeadamente, deve ter capacidade **de relacionar os dados provenientes de fontes heterogéneas**, mantendo uma ligação coerente entre elas.

Por outro lado, a **gestão de grandes fontes de conhecimento** será outro requisito importante para a implementação do *Miner*. Existe uma grande quantidade de dados em arquivos físicos, os quais devem ser geridos adequadamente. O CoSKS tem de

ter a capacidade de adquirir e indexar tais dados, de modo a torná-los disponíveis para o *Miner*. Por último, também é importante que os dados utilizados pelo *Miner* sejam fiáveis, de modo que as soluções encontradas para um determinado problema apresentem um grau de confiança considerável. Assim, um dos requisitos fundamentais para o correcto funcionamento do *Miner* é a **autenticidade** dos dados presente no CoSKS.

3.3 Bases Conceptuais

São três os conceitos fundamentais, nos quais este trabalho se centra: colaboração, conhecimento e raciocínio (figura 3.3).

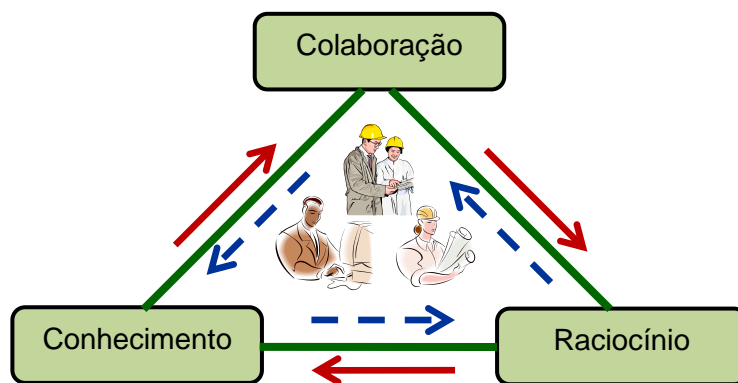


Figura 3.3: Bases conceptuais do trabalho (adaptado de [47])

A **colaboração** é um conceito fundamental em gestão do conhecimento pois permite gerar conhecimento e incentivar a criatividade dos indivíduos. Para a colaboração ser eficaz é necessário que todos os indivíduos no seio de um processo colaborativo, consigam lucrar com a partilha de experiências que são traduzidas em **conhecimento**. O **raciocínio** assume um papel fundamental, pois é responsável por traduzir os dados obtidos nos processos de colaboração, em informações efectivas. Por sua vez, as informações podem ser utilizadas na criação de novo conhecimento. De um modo geral, o raciocínio permite manipular os dados provenientes de diversas fontes independentes em colaboração umas com as outras e obter regras

lógicas e padrões comportamentais que possam de algum modo, melhorar o desempenho das empresas e evitar situações problemáticas. De facto, este tipo de informações pode constituir uma ferramenta valiosa para as empresas, contribuindo para a elaboração de planos de trabalho e estratégias inovadoras.

Fica assim claro que, para que a colaboração seja de facto uma mais-valia no processo de gestão do conhecimento, é necessário *aplicar* raciocínio sobre os dados recolhidos de modo a criar informação que poderá traduzir-se em conhecimento.

As empresas são responsáveis pela introdução de dados relativos a projectos que desenvolvem. Estes projectos são conduzidos através de um conjunto de reuniões, sendo cada uma destas reuniões considerada um **ponto de decisão**. Num ponto de decisão discutem-se questões e são tomadas decisões no sentido de encontrar soluções adequadas para as questões levantadas durante o desenvolvimento de um projecto (figura 3.4).

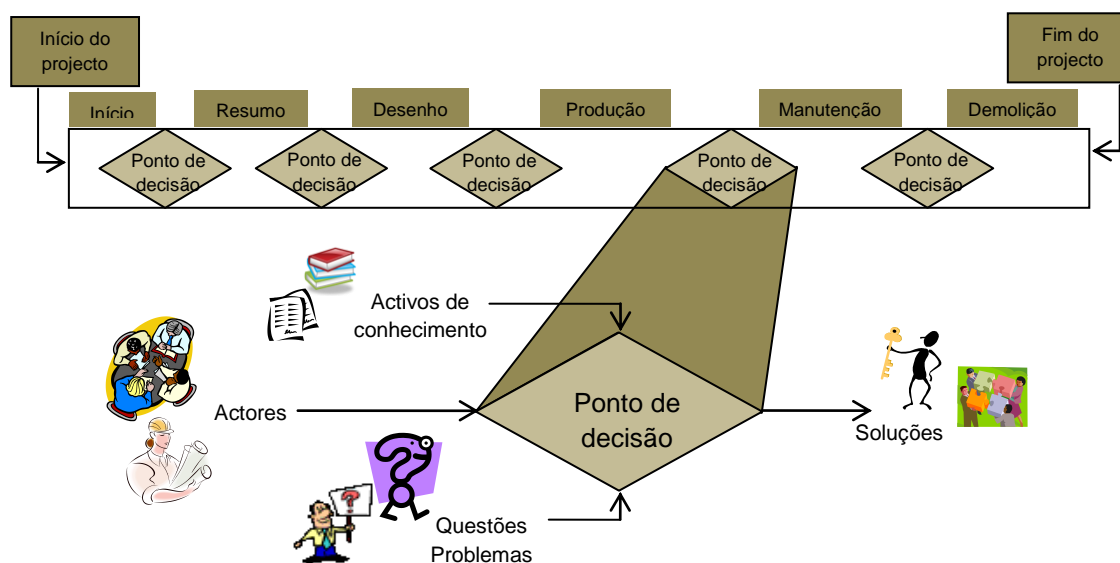


Figura 3.4: Ponto de decisões do projecto (adaptado de [48])

Eventualmente, as decisões podem traduzir-se num conjunto de tarefas a realizar. Cada ponto de decisão deve ser cuidadosamente planeado, através da elaboração prévia de agendas e os acontecimentos que ocorrem durante a reunião têm de ser

registados. Entre pontos de decisão existe uma monitorização permanente para avaliar a execução das tarefas definidas.

O **contexto** no qual o conhecimento é criado e usado influencia fortemente o processo. Deste modo, é necessário definir com precisão os contextos considerados neste trabalho. Isto porque cada profissional a trabalhar colaborativamente num determinado projecto, tem necessidades e visões próprias acerca das informações a usar. Por outro lado, diferentes tipos de projecto podem usar as mesmas informações de maneiras diferentes, nomeadamente pelo facto destas informações poderem ser tratadas de modo diferente dependendo de diversos factores, por exemplo, a fase do projecto ou o tipo do problema em discussão. Também, em diferentes tarefas as mesmas informações podem ter utilizações diferentes e é a questão a ser resolvida que define a relevância de uma informação.

O **elemento de conhecimento** constitui outro conceito básico adoptado neste trabalho. Representa *peças* de conhecimento que podem ser capturadas, armazenadas, publicadas, partilhadas e reutilizadas pelas equipas do projecto. Os elementos de conhecimento são de extrema relevância para o suporte da colaboração electrónica num determinado projecto. Alguns exemplos são: documentos comuns, projectos, soluções, agendas, tarefas e participantes (figura 3.5).

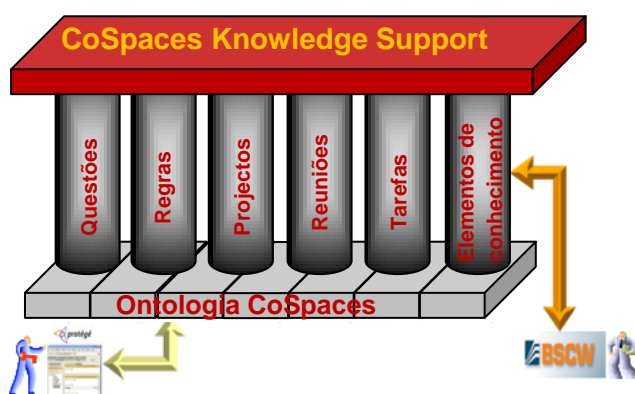


Figura 3.5: Níveis do conhecimento CoSpaces (adaptado de [47])

Os ficheiros, elementos de conhecimento comuns, são suportados e geridos por uma entidade conceptual que disponibiliza serviços para armazenamento, agrupamento, pesquisa e devolução de ficheiros.

3.4 DM como suporte à tomada de decisão

As sessões activas de colaboração de apoio à tomada de decisão seguem um fluxo de trabalho composto por quatro fases, nomeadamente, **fase de trabalho individual**, **fase de inicialização**, **fase de colaboração** e **fase de finalização/arquivo**, como mostra a figura 3.6.

A **fase de trabalho individual** constitui a identificação de um problema, sendo então elaborada uma descrição deste problema. Nesta fase, os indivíduos participantes no projecto colaboram assincronamente, sugerindo tarefas que possam resolver um problema.

A **fase de inicialização**, previamente à reunião, inclui a preparação da agenda e a selecção dos participantes da reunião. O responsável pela organização da reunião (e.g. gestor do projecto) pode identificar as principais questões a serem discutidas, assim como o melhor modo de as discutir.

A **fase de colaboração** consiste na reunião propriamente dita, na qual os participantes tentam chegar a um consenso relativamente às questões em agenda, utilizando os recursos apropriados. Esta fase também compreende o registo das decisões tomadas, incluindo novas tarefas e atribuição de responsabilidades. As decisões aqui tomadas devem ser suportadas pelo processo de tomada de decisão, de acordo com o contexto da questão discutida.

A **fase de finalização/arquivo**, também chamada pós-reunião, inclui a elaboração de minutas descrevendo o conjunto de conclusões baseadas nas decisões tomadas durante a reunião.

Sempre que um projecto é concluído, o gestor deve elaborar o *project post-mortem*, um documento que resume todos os indicadores relevantes (questões/decisões, desvios, participantes envolvidos, etc.). O *project post-mortem* representa um elemento de conhecimento que pode ser utilizado em projectos futuros, com o propósito de melhorar as performances, facilitar as boas práticas e eliminar erros conhecidos.

Cada uma das fases de uma sessão colaborativa contribui significativamente para o processo de tomada de decisão. Principalmente na fase de colaboração, o recurso a DM ganha importância e torna-se um elemento fundamental na tomada de decisão. Durante as reuniões, pelo facto de se disporem de dados manipulados por DM (existentes na base de dados e documentos de projectos anteriores), a informação relevante é identificada de uma forma rápida e ágil, facilitando o processo de tomada de decisão. Assim, para cada questão discutida durante a reunião, o recurso a DM pode ser a chave que permite encontrar a solução desejada de modo a resolver a questão.

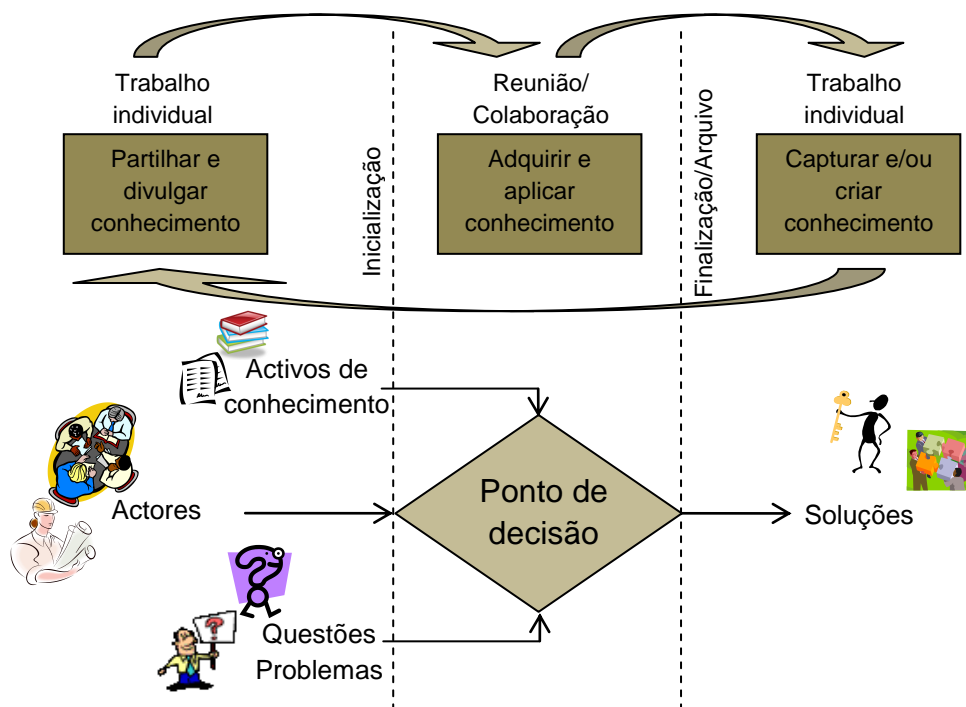


Figura 3.6: Fases do processo de colaboração (adaptado de [48])

3.5 Fundamentos técnicos do CoSKS

A estrutura CoSKS divide-se em quatro níveis: **apresentação**, **comportamento**, **serviços** e **conhecimento** (figura 3.7).

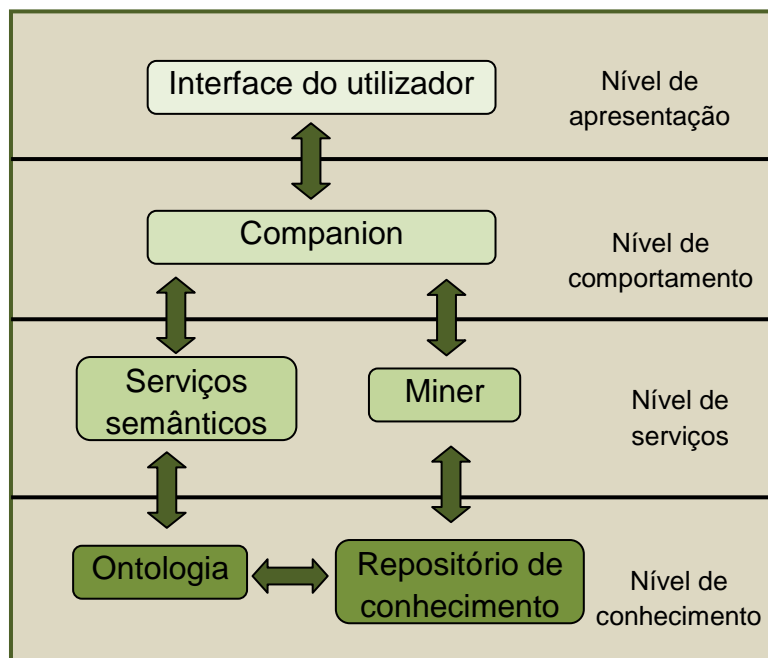


Figura 3.7: Níveis técnicos do CoSKS (adaptado de [48])

O **nível de apresentação** representa as interações do CoSKS com as aplicações externas, representadas pelo portal do CSF na figura 3.1. Estas interações possibilitam a colaboração entre as equipas de trabalho, envolvendo troca de conhecimento, actualização dos dados dos projectos, actores, tarefas, reuniões, etc.. Neste nível também ocorre a troca de ficheiros entre o CoSKS e uma entidade conceptual que gere os ficheiros e directorias, representada pelo BSCW na figura 3.1. Aqui são adicionados, removidos e actualizados documentos sobre os projectos, que podem ser consultados e eventualmente facilitar a resolução de problemas encontrados em projectos semelhantes.

No **nível de comportamento** encontra-se o *Companion* que é um componente proactivo, reactivo e autónomo responsável por alertar, monitorar e apoiar a tomada

de decisão em projectos. Algumas das funcionalidades oferecidas pelo *Companion* são:

- acompanhar constantemente o desenvolvimento de um projecto de modo a apresentar as melhores soluções para as questões/problemas que possam surgir;
- alertar para os prazos de execução de tarefas, desvios, datas de reuniões e respectivos participantes;
- responder a questões dos utilizadores; e
- coleccionar, organizar e filtrar informações e decisões; fazer um ponto da situação do projecto (prazos, custos).

O **nível de serviços** contém os serviços semânticos e o *Miner*. O *Miner* será abordado no próximo ponto do trabalho. Os serviços semânticos disponibilizam um conjunto de funcionalidades suportadas por uma ontologia. As funcionalidades principais são três: **(i)** filtrar e definir o contexto semântico; **(ii)** criar vectores semânticos; e **(iii)** indexar e pesquisar semanticamente.

A funcionalidade **filtrar e definir o contexto semântico** é responsável por representar cada elemento de conhecimento por um conjunto de palavras-chave, filtradas semanticamente de modo a definir o contexto semântico em que se enquadra. O contexto semântico constrói-se com base no tipo de dados envolvidos, por exemplo, tipo de projecto, fase de projecto, tarefa, problema, actor, etc..

Na segunda funcionalidade, **criar vectores semânticos**, cada elemento de conhecimento é definido por um vector semântico, o qual representa um conjunto de conceitos ontológicos que semanticamente o conseguem caracterizar. A atribuição de pesos semânticos aos conceitos ontológicos no vector semântico é efectuada de acordo com a relevância do elemento de conhecimento, num determinado contexto. O vector semântico é usado para classificar, comparar e

recuperar os elementos de conhecimento que permitem encontrar a melhor solução possível para determinado problema.

Em **indexar e pesquisar semanticamente**, os novos elementos de conhecimento são indexados semanticamente através da criação dos respectivos vectores semânticos. Quando uma nova questão/problema é introduzida no sistema, tendo em conta que a consulta é tratada pelo sistema como um elemento de conhecimento, o respectivo vector semântico é usado para procurar a semelhança semântica com outros vectores e dar uma resposta/solução para a questão/problema em curso.

Finalmente tem-se o **nível de conhecimento** composto pela ontologia e pelo repositório de conhecimento. A ontologia suporta o uso de expressões que contextualizam os elementos de conhecimento no repositório. A ontologia acrescenta um peso semântico nas relações entre os elementos de conhecimento armazenados. Cada conceito ontológico apresenta uma lista de termos equivalentes que podem ser usados para representá-lo semanticamente. Estes termos são tratados estatística e semanticamente de modo a criar o vector semântico dos elementos de conhecimento.

O repositório de conhecimento armazena todos os elementos de conhecimento usados no CoSKS, dos quais faz parte uma base de dados constituída pelos projectos, tarefas, actores, reuniões, questões, etc..

3.6 O Miner

O *Miner* é um componente que faz parte dos serviços avançados do CoSKS. Este contribui de certo modo para o processo de colaboração ao longo do desenvolvimento de um projecto. O *Miner* é responsável pelo processo de DM, também conhecido como extracção de informação de bases de dados, um processo de identificação de padrões válidos, existentes nos dados.

O *Miner* oferece funcionalidades para a descoberta de informações a partir de dados, presentes em base de dados ou documentos, que permitam responder a questões/problemas levantados durante a execução de um projecto. Esta constante procura de respostas permite não só resolver as diversas tarefas de um projecto em execução, mas sobretudo contribuir para um melhor desenvolvimento dos eventuais projectos futuros.

3.6.1 As interfaces do *Miner*

O *Miner* necessita de receber dados, trabalhá-los de modo a obter informação e transmiti-la para outro módulo avançado, responsável por tornar este conhecimento disponível ao utilizador, ou directamente ao utilizador.

A figura 3.8 mostra as interacções do *Miner* com os outros componentes do CoSKS.

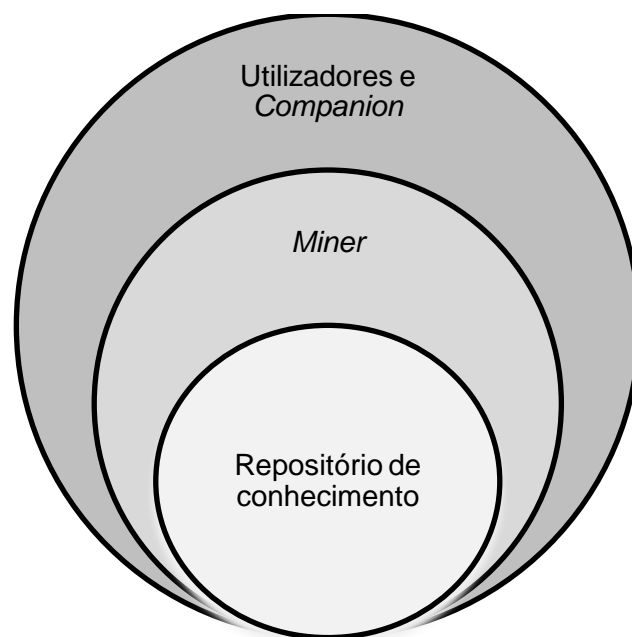


Figura 3.8: Interfaces do *Miner*

Basicamente o *Miner* interage com o *Companion* e utiliza o repositório de conhecimento.

Os serviços do *Miner* podem ser requisitados pelo *Companion*, sempre que necessário. Nestes casos, o *Miner* acede ao repositório de dados seleccionando os

dados adequados e sobre estes aplica um ou mais serviços de DM (figura 3.9). Os resultados são depois partilhados com o *Companion*, responsável por mostrá-los ao utilizador. Os resultados podem ser muito variados, consoante a tarefa requerida pelo utilizador e podem representar uma resposta a um problema que possa surgir ao longo do projecto, auxiliar nas escolhas a tomar durante a execução do projecto, etc..

Por outro lado, quando novos dados são adicionados ao repositório, geralmente à base de dados, o *Miner* entra em funcionamento. Por exemplo, quando a informação de um novo projecto, actor, tarefa é adicionada, existe a necessidade de agrupar os projectos, actores e tarefas com as características semelhantes.

3.6.2 Os serviços do *Miner*

De modo a dar resposta às necessidades que levaram à sua implementação, o *Miner* deve contemplar vários serviços (figura 3.9). Tal como anteriormente descrito, existem dois tipos de serviços distintos: os serviços executados quando o utilizador questiona o *Companion*, nomeadamente, **analisar dependência de dados, auxiliar tarefas, seleccionar actores, obter informação para questões, prever e avaliar desvios** e os serviços executados quando novos dados são introduzidos na base de dados dentro do repositório de conhecimento, que incluem: **caracterizar, classificar** e caracterizar por ***text mining***.

O serviço **caracterizar** permite especificar um conjunto de regras capazes de descrever as características de um conjunto específico de dados. Por exemplo, caracterizar um projecto com base no seu tempo de execução, custo ou outros dados específicos, assim como caracterizar um actor, com base no número de participações em projectos e profissão. Este serviço pode ser aplicado a projectos, tarefas, questões, actores, etc..

O serviço **obter informação para questões** inclui a pesquisa nos dados pela resposta a uma determinada questão colocada pelos utilizadores. Por exemplo, *há necessidade de alargar o prazo de execução do projecto, qual o impacto no custo do projecto? O projecto está atrasado, quantos trabalhadores são necessários adicionar ao projecto de modo a cumprir os prazos?*

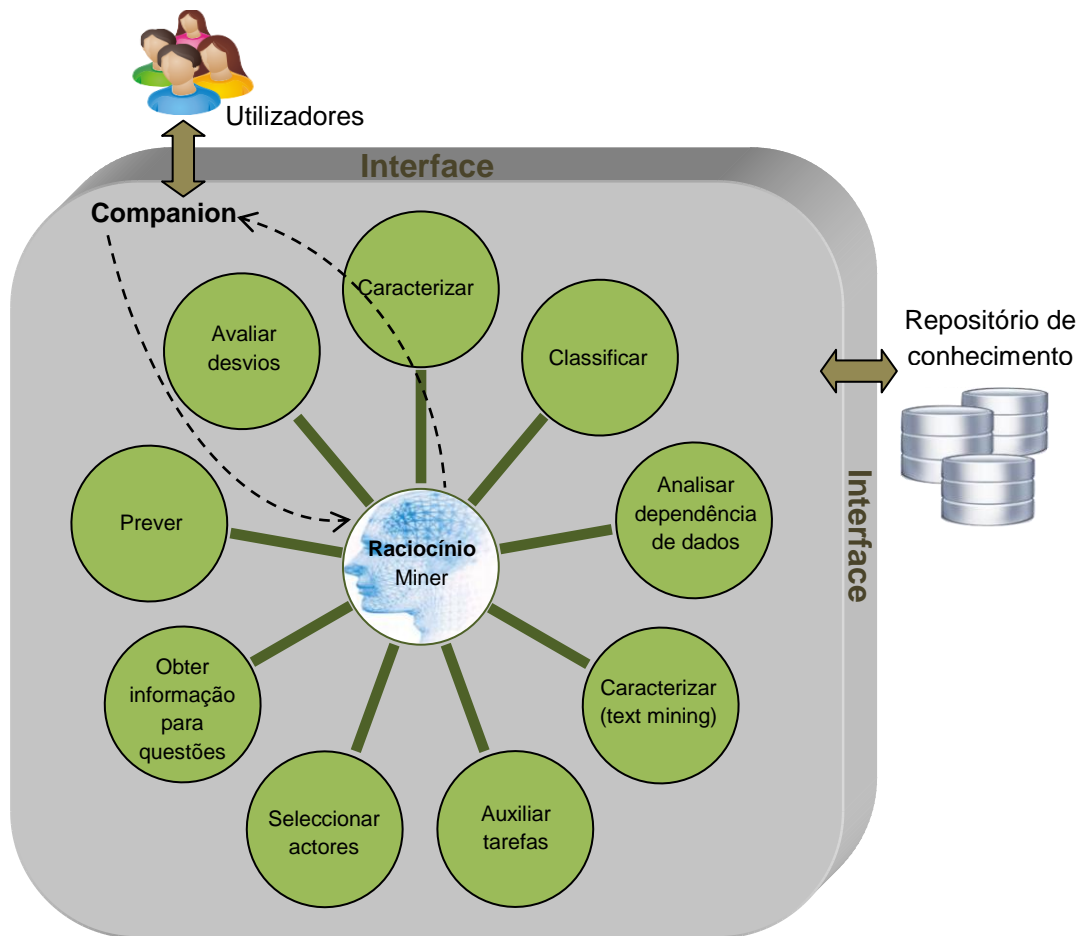


Figura 3.9: Serviços do *Miner*

No serviço **classificar** utilizam-se conjuntos de dados de treino (ou seja, conjuntos de dados cuja classe é conhecida) para construir modelos para cada classe, baseados nas características desses dados. Criam-se assim conjuntos de regras de classificação, que podem ser usados para classificar os dados futuros e desenvolver uma melhor compreensão de cada classe no conjunto de dados. Por exemplo, quando é iniciado um novo projecto (com as suas respectivas características) pode

ser útil classificá-lo, de modo a compreender melhor os problemas que possam surgir ao longo do desenvolvimento do mesmo. Isto é válido também para as tarefas, questões e actores, de tal modo que todas estas classificações podem contribuir para um melhor desenvolvimento do projecto (nomeadamente, encurtar o tempo da realização das tarefas com base na experiência de tarefas anteriores, que melhor se enquadram no projecto).

O serviço **analisar dependência de dados** permite efectuar uma pesquisa com recurso a regras de associação (na forma $A_1 \wedge \dots \wedge A_i \rightarrow B_1 \wedge \dots \wedge B_j$), tentando encontrar dependências entre os dados. A ideia básica é identificar na base de dados, situações em que a ocorrência de um atributo específico implica a ocorrência de outro atributo com alguma probabilidade. No contexto deste trabalho, um exemplo de uma regra de associação pode ser *a descoberta de que para projectos, com tempos de execução previstos para 3 anos e custos de 100 mil euros, em 70% dos casos o tempo de execução do projecto não é cumprido*. A descoberta de informação relevante com base em regras de associação levanta questões sobre o correcto planeamento dos projectos e o conhecimento destas regras pode evitar problemas futuros. Este serviço deve ser aplicado a todos os dados existentes na base de dados, pois a determinação da dependência entre os dados permite compreender melhor os futuros projectos.

Em **caracterizar documentos** pretende-se descrever os documentos que vão sendo criados, usados e adicionados na base de dados ao longo da execução dos projectos, reuniões e tarefas. Uma boa caracterização dos documentos pode reduzir o tempo de procura de soluções, na forma de documentos capazes de auxiliar na resolução de tarefas. Este serviço consiste em adicionar, automaticamente, alguns campos à base de dados sempre que um novo documento é criado e adicionado,

por exemplo, o nome do autor, palavras-chave que descrevam o conteúdo do documento, datas, etc...

No serviço **auxiliar tarefas** procuram-se soluções capazes de ajudar na resolução de tarefas. Neste serviço o *Companion* comunica com o *Miner* usando palavras-chave que possam caracterizar a tarefa. Por sua vez, o *Miner* pesquisa nos dados e tenta encontrar informações em documentos e tarefas, com características semelhantes às pedidas, devolvendo ao *Companion* os dados que encontrou.

O serviço **seleccionar actores** procura por um conjunto de especialistas que se enquadrem no contexto de um novo projecto. Isto é, com base nas características do novo projecto (tempo de execução, tipo de projecto, descrição, etc.) o serviço determina que tipos de actores são mais apropriados.

O serviço **prever** é responsável por elaborar previsões sobre possíveis valores de dados em falta assim como a distribuição de valores de determinados atributos num conjunto de dados. Este serviço envolve a descoberta de um conjunto de atributos relevantes para o atributo de interesse e a previsão do seu valor, com base no conjunto de dados semelhantes ao objecto seleccionado. No caso deste trabalho, este serviço pode ser útil em algumas situações, por exemplo, a previsão do número de dias de execução de um projecto, pode ser realizada com base na média dos dias de execução de projectos semelhantes anteriores. De facto, existem muitas previsões possíveis, seleccionando o conjunto de atributos que, de algum modo, contribuam para o dado a prever, sendo que a previsão é sempre feita com base nos dados existentes, ou seja, com base no histórico dos projectos.

Por fim, o serviço **avaliar desvios** tem como objectivo determinar as alterações que se observam durante a execução de um projecto, tais como o cumprimento dos prazos e custos.

4. Implementação

Este capítulo descreve as tecnologias adoptadas no presente trabalho assim como o processo de implementação do *Miner*. De modo a compreender as funcionalidades do *Miner*, serão apresentadas a formalização funcional e a visão estrutural da infra-estrutura do software *Miner*. Para descrever a visão funcional recorrem-se a diagramas UML (Unified Modeling Language), onde se identificam os tipos de dados sobre os quais cada serviço *Miner* actua de modo a produzir informação. Por sua vez, na descrição da visão estrutural utiliza-se a metodologia ICE (Interface, Controlo e Entidade).

4.1 Tecnologias adoptadas no trabalho

Na tabela 4.1 descrevem-se as várias tecnologias adoptadas, referindo a sua utilização no desenvolvimento do *Miner*.

Tabela 4.1: Tecnologias adoptadas

Tecnologia	Definição	Utilização
NetBeans	Ambiente de desenvolvimento integrado disponível para Windows, Mac, Linux e Solaris. Trata-se de uma ferramenta gratuita adequada para situações em que se pretende desenvolver software em ambiente Web, desktop e aplicações móveis com a plataforma Java, C/C++, PHP, JavaScript, Groovy e Ruby [49].	Esta ferramenta serve de suporte ao desenvolvimento do software, assim como de toda a programação Java, JSP (Java Server Pages) e HTML (HyperText Markup Language).
JSP	Tecnologia Java que permite o desenvolvimento rápido de aplicações baseadas na Web, independentes da plataforma [50].	A tecnologia JSP possibilita a criação de páginas Web dinâmicas, com HTML ou XML (Extensible Markup Language).

Hibernate	Biblioteca ORM (Object Relational Mapping) disponível para linguagem Java, que facilita o mapeamento de classes Java para tabelas da base de dados e também de tipos de dados Java para dados SQL (Structured Query Language) [51].	O uso do Hibernate permite reduzir o tempo de programação dispendido em tarefas comuns, eliminando a necessidade de processamento manual de dados usando SQL e JDBC (Java Database Connectivity).
Tomcat	Servidor de aplicações Web Java que permite implementar as tecnologias JSP e Java Servlet, proporcionando um ambiente de servidor Web adequado para a execução do código Java [52].	Servidor Web utilizado para executar a aplicação, ou seja as páginas Web.
Visual Paradigm for UML	Tecnologia de apoio ao desenvolvimento de software, permitindo aos utilizadores o desenho, integração e implementação das suas aplicações de uma forma fácil, rápida e de qualidade. O Visual Paradigm suporta diferentes tipos de diagramas UML, nomeadamente, diagramas de classes, de casos de uso, de sequência, de implementação, etc., [53].	Utilizado para desenhar os diagramas UML e diagramas de classe que facilitam a compreensão do desenvolvimento do software. Também possibilita a criação de bases de dados gerando automaticamente o respectivo código SQL, assim como gerar e mapear código Java, reduzindo o tempo de programação de código repetitivo.
MySQL	Sistema para gestão de bases de dados que utiliza a linguagem SQL. Trata-se de uma aplicação de fácil utilização e de excelente desempenho e estabilidade [54].	Responsável pelo armazenamento dos dados dos projectos criados ao longo do tempo.

Struts 2	Estrutura para a criação de aplicações Web Java que adopta um modelo de arquitectura MVC (Model View Controller) para otimizar o ciclo de desenvolvimento das páginas JSP, desde a construção, implementação e manutenção das aplicações [55].	O recurso ao Struts 2 permitiu estruturar a programação das páginas JSP, separando o código HTML e Java das acções (ActionServlet) realizadas em cada página.
RapidMiner	Software gratuito que oferece soluções e serviços nas áreas de análise preditiva, DM e TM. O RapidMiner concentra-se em análises automáticas de dados estruturados (e.g. base de dados) e dados não estruturados (e.g. documentos de texto) [43].	Responsável pela aplicação de técnicas DM e TM sobre os dados dos projectos.
BSCW	Ferramenta para colaboração entre grupos, eficiente e de fácil utilização que permite a partilha de ficheiros. O BSCW possibilita a criação de documentos, contactos, tarefas, notas ou compromissos num ambiente de trabalho compartilhado [56].	Utilizado para armazenar os documentos criados e consultados nos projectos pelos utilizadores.

A figura 4.1 representa as ligações entre as várias tecnologias utilizadas. O desenvolvimento do software é suportado pelo **NetBeans**. Esta ferramenta inclui *frameworks* opcionais das quais foram seleccionadas o **Hibernate** e o **Struts2**. O Hibernate realiza o mapeamento entre as classes existentes no NetBeans e os dados existentes no **MySQL**, enquanto o Struts 2 permite estruturar a arquitectura num modelo MVC para otimizar o ciclo de desenvolvimento das páginas **JSP**. O **Tomcat** permite executar a aplicação, isto é, permite mostrar a interface CoSKS aos utilizadores. Por sua vez, a tecnologia **BSCW** interage com a interface CoSKS de modo a armazenar dados sob a forma de documentos, permitindo depois a utilização destes por parte do *Miner*. Por último, o **RapidMiner** interage com o

MySQL de forma a aplicar técnicas de DM sobre os dados. A importância destas duas últimas tecnologias na implementação do *Miner* será discutida posteriormente neste capítulo.

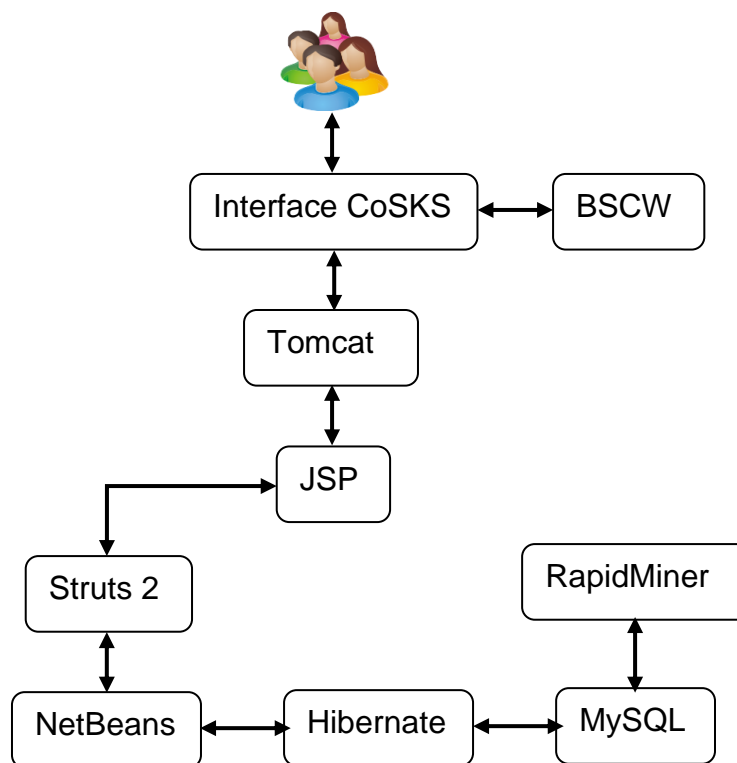


Figura 4.1: Interações entre as tecnologias adoptadas

4.2 A infra-estrutura do software *Miner*

Neste trabalho adoptaram-se diagramas UML e a metodologia ICE para implementar a infra-estrutura do software *Miner*.

UML é uma linguagem padrão orientada a objectos, que permite modelar formalmente qualquer sistema (e.g. sistema de *software*), facilitando a comunicação com outras aplicações e tornando a modelação do sistema mais compreensível. Deste modo, a UML foi utilizada para estruturar, visualizar informação e modelar os processos ao longo do desenvolvimento do sistema, na forma de diagramas de casos de uso e diagramas de classes, possibilitando a representação das várias funcionalidades do *Miner*.

Para fornecer uma visão estrutural do *Miner* recorreram-se a diagramas de classes. Estes diagramas permitem representar formalmente a estrutura de um sistema através da utilização de classes, onde estas são simplesmente uma descrição da funcionalidade dos objectos onde partilham as mesmas propriedades (semântica, atributos, métodos e relações). Neste trabalho, adoptou-se a metodologia ICE (Interface Controlo e Entidade) para dividir as classes na qual a estrutura do sistema se baseia.

4.2.1 Formalização do *Miner*

Para uma melhor compreensão das funcionalidades do *Miner*, foi elaborado um conjunto de diagramas de casos de uso UML, especificando as relações entre tipos de dados e os serviços que lhes estão associados.

Note-se que, num determinado projecto, os intervenientes podem ser classificados como gestores ou participantes, sendo que esta classificação se traduz em diferentes responsabilidades e acessibilidades. Os gestores têm uma maior liberdade de manipulação de dados, o que se reflecte também ao nível da aplicação de serviços *Miner*, de acordo com o especificado nos diagramas UML.

A figura 4.2 mostra os serviços *Miner* disponíveis para a análise de projectos, assim como as fases do projecto em que podem ser aplicados. O serviço **caracterizar** selecciona um conjunto de atributos, tais como custo, datas, descrição, que permite descrever o projecto de tal modo que o serviço **classificar**, tendo por base o valor destes atributos, tem a capacidade de enquadrar o projecto num determinado contexto, associando-o a um determinado grupo de projectos predefinidos. Pelo facto de possibilitar a organização dos projectos em grupos específicos, estes serviços são úteis em situações de pesquisa, particularmente por permitirem restringir a pesquisa a um determinado grupo de projectos.

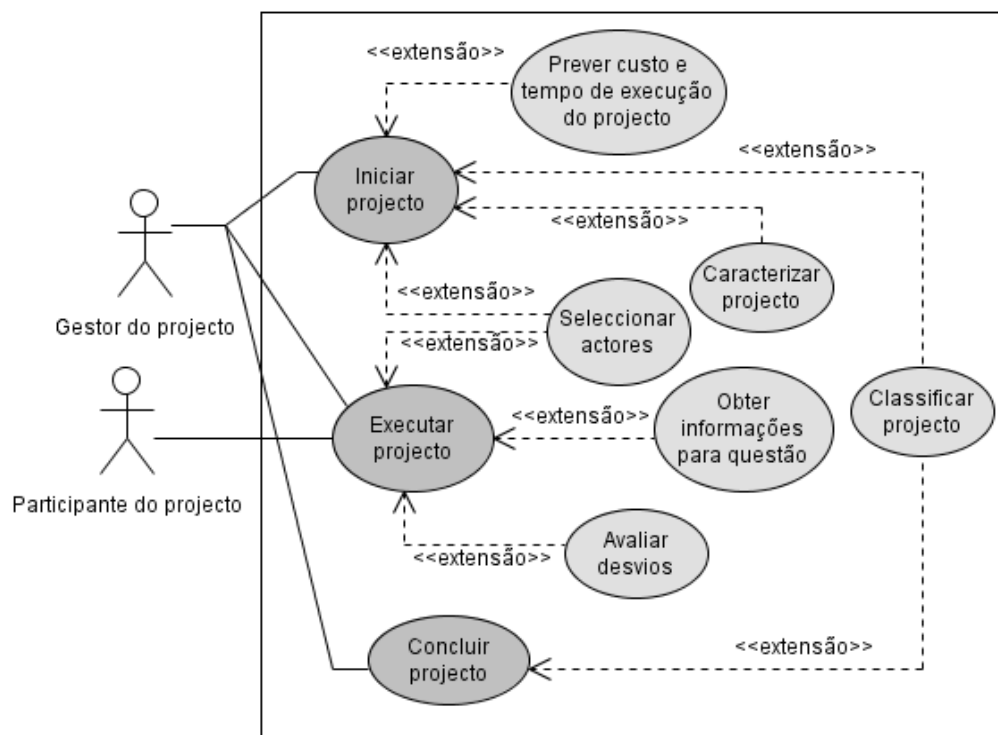


Figura 4.2: Diagrama de casos de uso dos projectos

O serviço **prever** estima o custo ou tempo de execução do projecto e pode ser utilizado na fase de início de projecto. Recorrendo à classificação em grupos, resultante da aplicação dos serviços caracterizar e classificar, este serviço permite determinar uma estimativa do custo ou tempo de execução do projecto, baseando-se nos dados relativos a projectos do mesmo grupo.

Durante as fases de iniciação ou execução de um projecto, pode-se recorrer ao serviço **seleccionar actores** de modo a identificar os participantes mais adequados para o projecto em questão. Esta selecção de actores é realizada de acordo com a frequência de participações que cada actor apresenta no histórico de projectos semelhantes.

O serviço **obter informações para questões** tem como objectivo a pesquisa de questões equivalentes a uma determinada questão em curso. A identificação destas questões equivalentes, assim como documentos e série de tarefas associados, podem promover informação útil para a resolução da questão em curso.

Por último, o serviço **avaliar desvios** facilita o reconhecimento de alterações ao planeamento inicial do projecto, particularmente no que respeita a prazos ultrapassados e gastos excedidos.

A figura 4.3 mostra o contexto em que os serviços podem ser aplicados na análise das reuniões. De modo análogo ao previamente descrito para os projectos, os serviços **caracterizar** e **classificar** podem ser aplicados com o mesmo propósito para o estudo das reuniões (assim como das tarefas e questões).

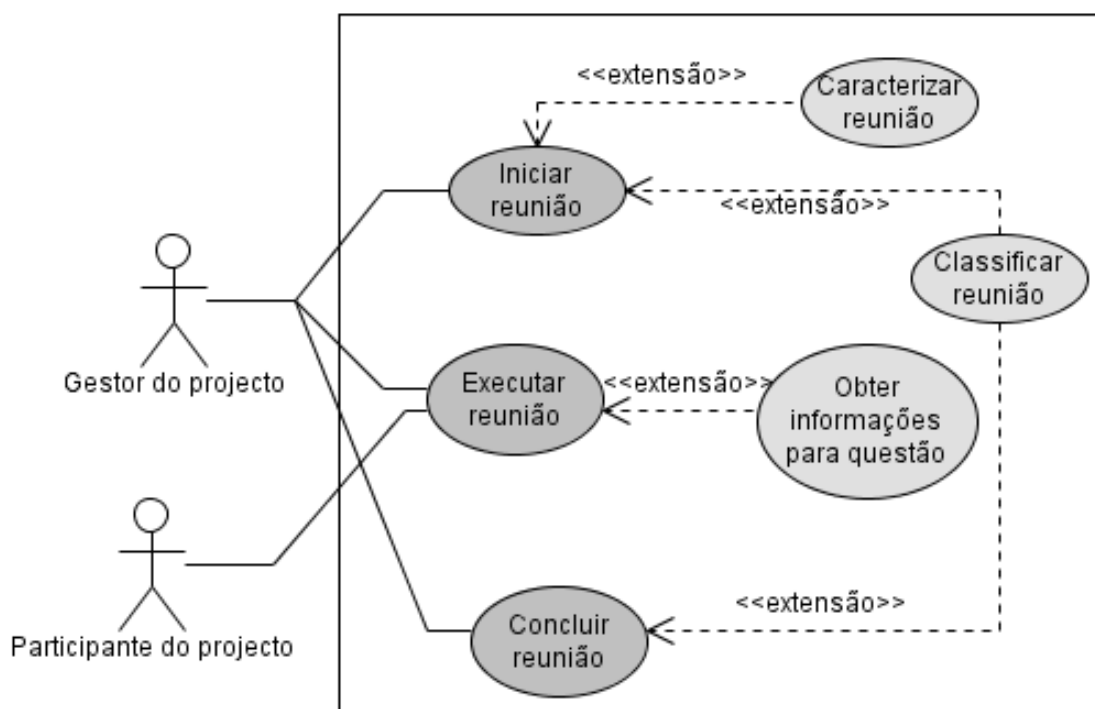


Figura 4.3: Diagrama de casos de uso das reuniões

Para além dos serviços **caracterizar** e **classificar** é possível utilizar também o **serviço obter informações para questões**. Com este serviço pretende-se obter informação que de algum modo se relacione com as questões/problemas a discutir na reunião e facilite a análise e/ou resolução dos mesmos. A informação que é aqui mencionada pode consistir em vários documentos referentes à questão/problema em estudo, ou em descrições de tarefas previamente desempenhadas na resolução de situações semelhantes. De notar que este serviço não permite a resposta directa

às questões em discussão, apenas faculta informação que pode contribuir para a elaboração de estratégias a adoptar.

Na figura 4.4 apresentam-se os serviços disponíveis para a análise de tarefas. Alguns destes serviços são semelhantes aos descritos anteriormente, no estudo das reuniões e projectos. Por exemplo, no serviço **seleccionar actores** procede-se à identificação dos participantes mais adequados para a realização das tarefas em questão tendo por base a sua participação em tarefas anteriores.

No que respeita à utilização do serviço **avaliar desvios**, pretende-se monitorizar a execução da tarefa, avaliando se está a decorrer de acordo com o previsto.

O serviço **auxiliar tarefa** pretende descrever uma lista de tarefas semelhantes à tarefa em análise, de modo a facilitar o planeamento da nova tarefa.

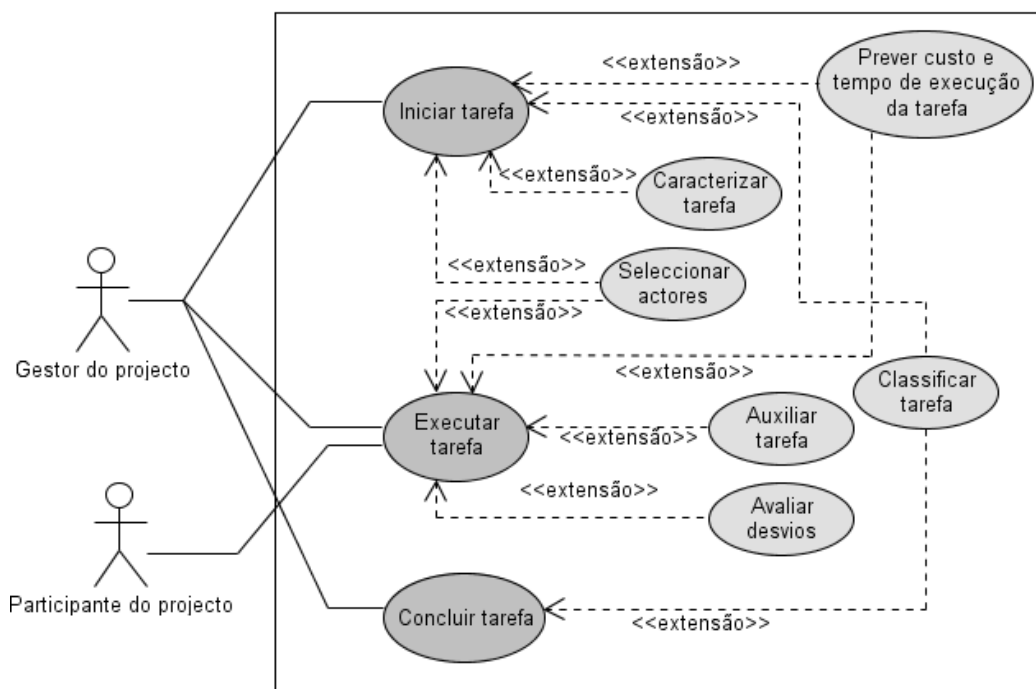


Figura 4.4: Diagrama de casos de uso das tarefas

Na figura 4.5 são apresentados os serviços que podem ser utilizados na análise das questões que surgem durante o desenvolvimento do projecto. Os serviços **caracterizar**, **classificar** e **obter informações para a questão** podem ser aqui aplicados com o propósito de identificar questões com características semelhantes à

questão em análise. Uma vez identificadas estas questões, procede-se à selecção de informações relevantes, nomeadamente sobre o tipo de tarefas realizadas e respectivos participantes ou documentos consultados.

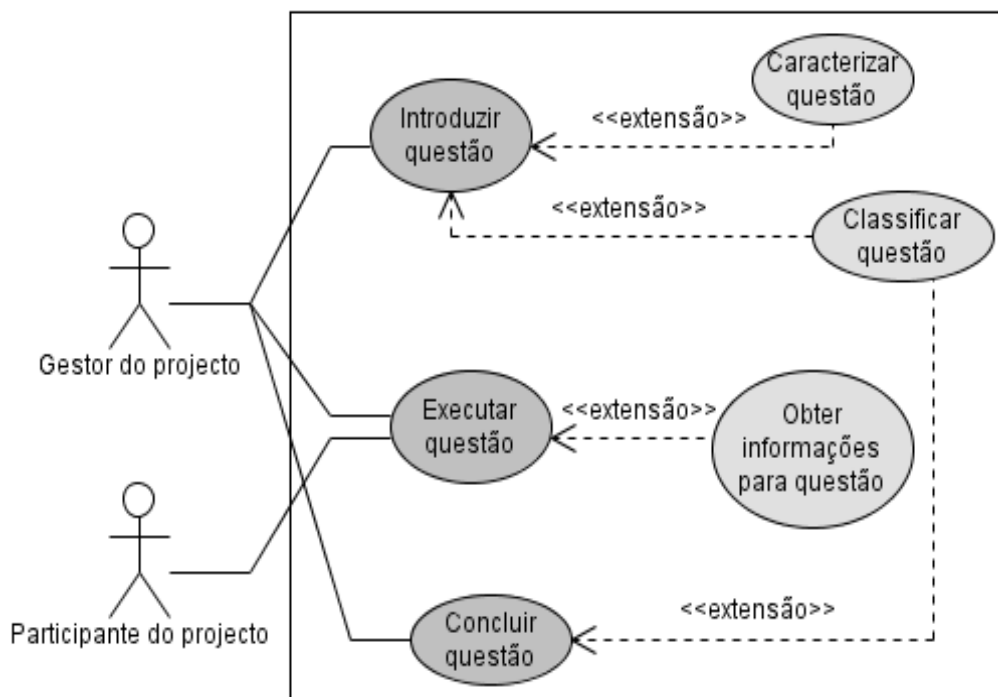


Figura 4.5: Diagrama de casos de uso das questões

A figura 4.6 representa os serviços aplicados no estudo dos documentos criados e/ou actualizados durante o desenvolvimento de um projecto.

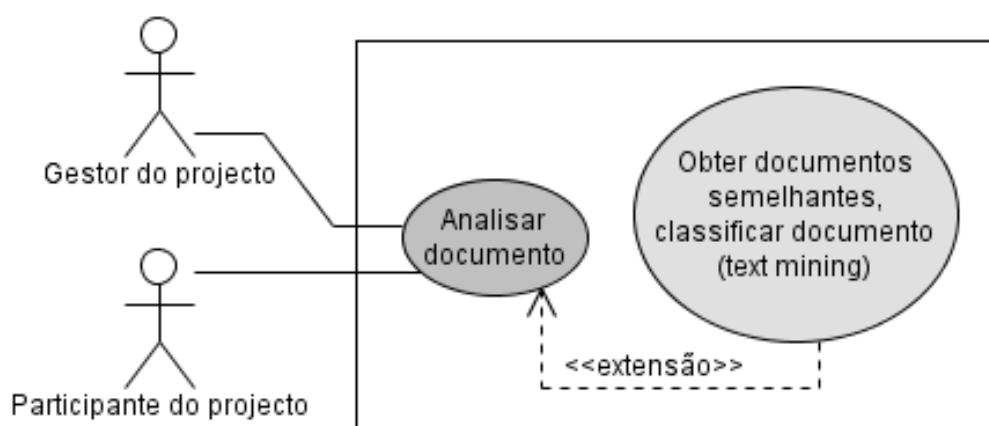


Figura 4.6: Diagrama de casos de uso dos documentos

Sempre que um documento é inserido ou actualizado, pela utilização do serviço **classificar documento**, são identificadas expressões presentes no documento

incluindo nomes, datas e outras palavras-chave, que caracterizam o conteúdo do documento. Com esta classificação, os documentos importantes para a resolução de alguma das situações previamente descritas são mais facilmente identificados.

Todos os resultados obtidos durante a execução dos serviços *Miner* são mostrados ao gestor do projecto e participante, sempre que possível, na forma de tabelas. Os resultados obtidos pela utilização dos serviços *Miner* devem contribuir para a tomada de decisões durante o desenvolvimento de um projecto. Contudo, são os indivíduos intervenientes no projecto que realizam uma análise crítica das informações disponibilizadas pelo *Miner*, elaborando as estratégias a adoptar, sendo os gestores do projecto os responsáveis pela aprovação das decisões a tomar.

4.2.2 Visão estrutural do *Miner*

Neste trabalho adoptou-se a metodologia ICE, a qual está dividida nos níveis **interface, controlo e entidade**, onde:

- Interface: contém todas as classes que possibilitam as interacções do sistema com o exterior;
- Controlo: estabelece a ligação entre os níveis interface e entidade, processando os pedidos a partir da interface e utilizando os dados existentes na entidade; e
- Entidade: inclui todas as classes que realizam a gestão dos dados utilizados pelo sistema, podendo também ser designada por base de dados.

Na figura 4.7 encontra-se representada a arquitectura resultante do uso da metodologia ICE, com informação das classes mais importantes para a implementação das funcionalidades do *Miner*, que estão associadas a cada um dos níveis anteriormente descritos.

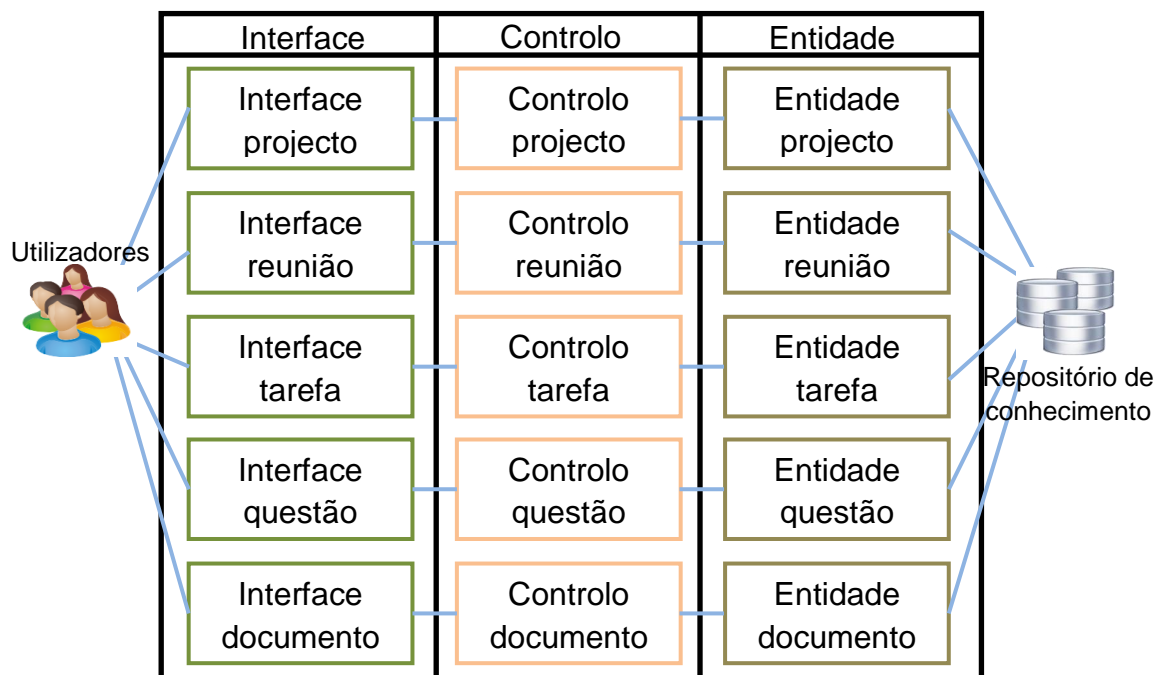


Figura 4.7: Diagrama de classes ICE

4.2.2.1 O nível interface

O nível interface contém todas as classes que gerem as interações dos utilizadores com o sistema. Estas classes tornam possível aos utilizadores partilhar, adicionar, pesquisar, actualizar e apagar dados do sistema.

A **interface projecto** (figura 4.8) é uma classe que possibilita ao utilizador introduzir, alterar, remover ou consultar dados sobre projectos.

Interface project
-int idProject
+insertProject() +consultProject() +updateProject() +insertFamily() +insertAttribute() +addActorProject() +listPredictionProject() +listSimilarProjects() +listRankingActorsProject()

Figura 4.8: Interface projecto

No que respeita ao funcionamento do *Miner*, esta classe é importante porque permite ao utilizador requisitar e receber informação relativa a projectos. É portanto,

esta interface a responsável por apresentar ao utilizador eventuais resultados da utilização de serviços associados ao estudo de projectos, nomeadamente o serviço **caracterizar**, **classificar**, **prever** custo e tempo de execução do projecto e **seleccionar actores** que se enquadrem no contexto do projecto.

De modo análogo à interface do projecto, as **interfaces reunião** (figura 4.9), **questão** (figura 4.10) e **tarefa** (figura 4.11), permitem ao utilizador introduzir, alterar, remover ou consultar dados sobre reuniões, questões ou tarefas, respectivamente. Por exemplo, no caso da interface tarefa, esta classe pode ser utilizada para **seleccionar actores** para determinadas tarefas, assim como informar o utilizador sobre alguns desvios que estejam a ocorrer.

Interface meeting
-int idMeeting
+insertMeeting() +updateMeeting() +removeMeeting() +consultMeeting() +listSimilarIssuesForMeeting()

Figura 4.9: Interface reunião

Interface issue
-int idIssue
+insertIssue() +removeIssue() +updateIssue() +consultIssue() +listSimilarIssue()

Figura 4.10: Interface questão

Interface task
-idTask
+insertTask() +removeTask() +updateTask() +consultTask() +listSimilarTasks() +listPredictionTask() +listRankingActorsTask()

Figura 4.11: Interface tarefa

A última classe do nível interface aqui referida é a **interface documento**, representada na figura 4.12. Esta classe permite ao utilizador adicionar, actualizar e remover um documento associada a um projecto, tarefa ou questão e possui algumas funções responsáveis pela aplicação do serviço **caracterizar (TM)**.

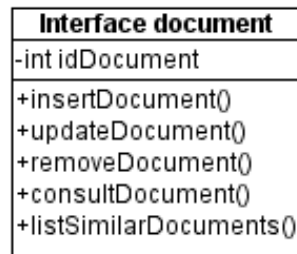


Figura 4.12: Interface documento

4.2.2.2 O nível controlo

O nível controlo estabelece a ligação entre o nível interface e o nível entidade. Após a introdução de um pedido no nível interface, por parte do utilizador, o nível controlo é responsável por transmitir às classes do nível entidade quais os dados necessários para executar o pedido do utilizador. Em seguida, o nível entidade fornece ao nível controlo, os dados seleccionados e as classes deste nível processam os dados de modo a obter os resultados pretendidos, que são então enviados para o nível interface, ficando disponíveis para consulta do utilizador.

Este nível é constituído pelas classes **controlo projecto** (figura 4.13), **controlo reunião** (figura 4.14), **controlo tarefa** (figura 4.15), **controlo questão** (figura 4.16) e **controlo documento** (figura 4.17), responsáveis por processar os dados recebidos ou enviados do nível interface e enviá-los para o nível entidade e vice-versa, de modo a executar os pedidos do utilizador no que diz respeito a projectos, reuniões, tarefas, questões ou documentos, respectivamente. Tal como as classes da interface, as classes do controlo possuem funções importantes na aplicação de serviços **prever**, **auxiliar tarefa**, **obter informação para questões**, **caracterizar**, **classificar**, etc..

Control project
-int idProject
+insertProject(parameters) +updateProject(int idProject, parameters) +consultProject(int idProject) +classificationProject(int idProject) +addActorProject(int idProject, int idActor) +predictionProject(int idFamily) +similarProjects(String keywords [], int idFamily) +rankingActorsProject(int idFamily)

Figura 4.13: Controlo projecto

Control Meeting
+insertMeetings(parameters) +updateMeetings(int idMeeting, parameters) +removeMeetings(int idMeeting) +consultMeeting(int idMeeting) +similarIssuesForMeeting(String keywords [])

Figura 4.14: Controlo reunião

Control issue
-int idIssue
+insertIssue(parameters) +updateIssue(int idIssue, parameters) +remove(int idIssue) +consultIssue(int idIssue) +classificationIssue(int idIssue) +similarIssues(String keywords [])

Figura 4.15: Controlo questão

Control task
-int idTask
+insertTask(parameters) +updateTask(int idTask, parameters) +removeTask(int idTask) +classificationTask(int idTask) +addActorToTask(int idTask, int idActor) +similarTasks(String keywords []) +predictionTask(String keywords []) +rankingActorsTask()

Figura 4.16: Controlo tarefa

Control document
-int idDocument
+insertDocument(parameters) +updateDocument(int idDocument, parameters) +removeDocument(int idDocument) +consultDocument(int idDocument) +similarDocuments(String keywords [])

Figura 4.17: Controlo documento

4.2.2.3 O nível entidade

Obviamente o *Miner* necessita de dados para a execução dos seus serviços. Os dados utilizados encontram-se armazenados numa base de dados de projectos gerida pela ferramenta MySQL. Por outro lado, o *Miner* também utiliza dados na forma de documentos, sendo que a tecnologia BSCW é a responsável pelo seu armazenamento.

Deste modo, o *Miner* tem de aceder aos repositórios de dados para efectuar os seus serviços. O nível entidade é o responsável por comunicar com o repositório promovendo o acesso à base de dados, que adopta um modelo de diagrama entidade relação (DER) (apresentado no **Anexo A**) e ao BSCW para obter os dados necessários. Na posse destes dados, o *Miner* realiza os seus serviços com o objectivo de produzir informação.

O nível entidade partilha a estrutura dos níveis anteriores, sendo constituído pelas classes **entidade projecto**, **entidade reunião**, **entidade questões**, **entidade tarefa** e **entidade documento**.

A figura 4.18 apresenta as principais funções da **entidade projecto** responsáveis pela gestão de dados nas tabelas relacionadas com projectos, presentes no repositório de dados. Estas incluem as tabelas projectos, extra projectos, família de projectos, actores dos projectos e atributos das famílias dos projectos, representadas na figura 4.19.

Entity project
+insertProject() +updateProject() +consultProject() +insertFamily() +insertAttribute() +parametersPredictionProjects() +parametersSimilarProjects() +parametersRankingActorsProject()

Figura 4.18: Entidade projecto

As tabelas da figura 4.19 armazenam os dados que podem ser utilizados pelo *Miner* na implementação dos serviços **prever**, **seleccionar actores**, **caracterizar** e **classificar** projectos.

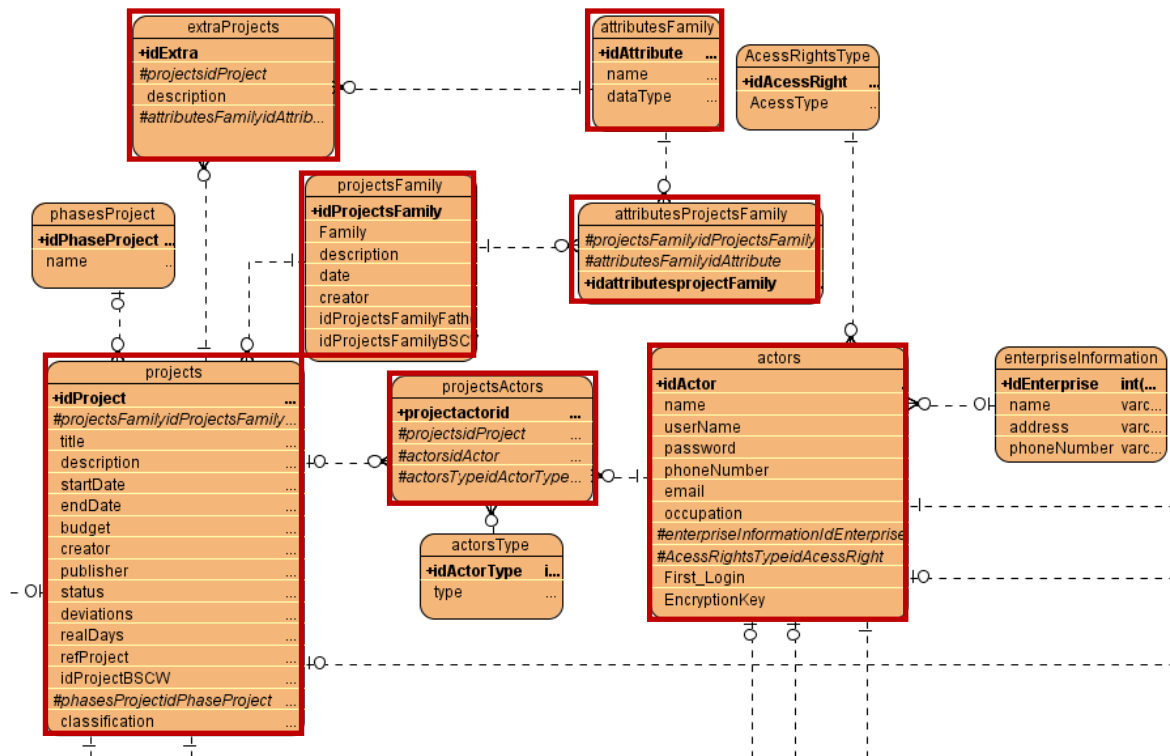


Figura 4.19: Excerto do DER (Ligações dos projectos)

A **entidade reunião**, na figura 4.20, contém funções que permitem gerir os dados presentes nas tabelas relacionadas com reuniões.

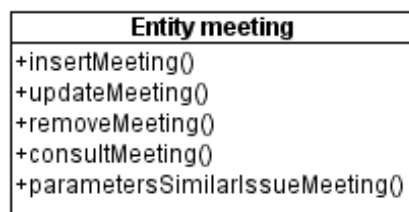


Figura 4.20: Entidade reunião

As tabelas afectadas por esta entidade encontram-se representadas no excerto do DER da figura 4.21. No que diz respeito aos serviços, esta classe possui a função responsável por transferir os atributos necessários para a aplicação do serviço **obter informação para questões**, que estejam a ocorrer numa reunião.

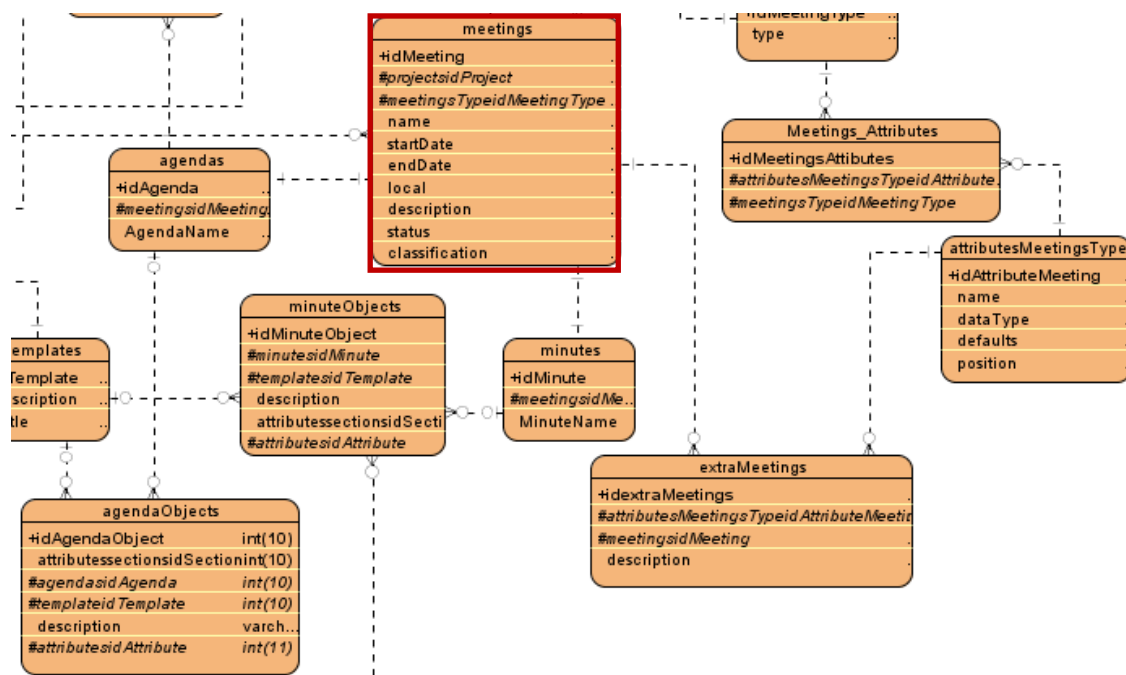


Figura 4.21: Excerto do DER (Ligações das reuniões)

A **entidade questão** (figura 4.22) e **entidade tarefa** (figura 4.23) permitem gerir os dados sobre questões e tarefas que são armazenados nas tabelas presentes no repositório de dados e que se encontram representadas na figura 4.24.

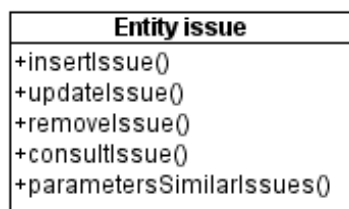


Figura 4.22: Entidade questão

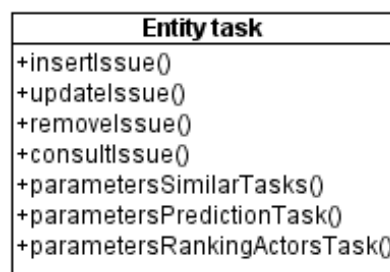


Figura 4.23: Entidade tarefa

A tabela que armazena dados relativos a tarefas, representada na figura 4.24 com a denominação *Tasks*, permite fazer a distinção entre dois tipos de tarefas: tarefas

associadas a questões que são realizadas para solucioná-las ou tarefas relacionadas com situações do dia-a-dia do projecto.

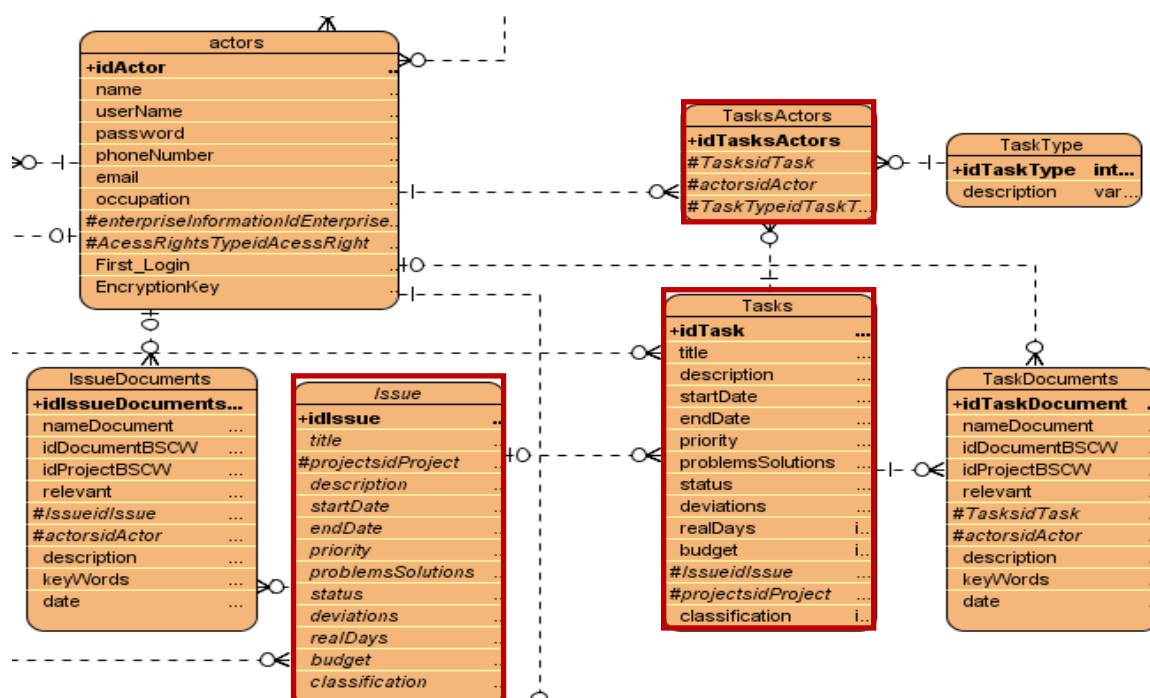


Figura 4.24: Excerto do DER (Ligações das tarefas e questões)

Por último, a **entidade documento** (figura 4.25) inclui todas as funções importantes nas acções de introdução, actualização e consulta de documentos utilizados durante o desenvolvimento de projectos, resolução de questões e execução de tarefas.

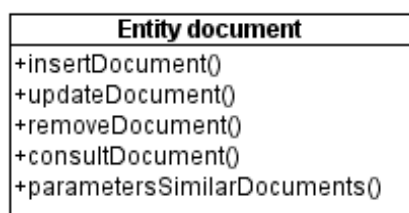


Figura 4.25: Entidade documento

As tabelas presentes no repositório apenas armazenam informações chave sobre estes documentos (figura 4.26), sendo que o conteúdo físico dos documentos é arquivado no BSCW. A pesquisa de documentos armazenados nestas tabelas é efectuada pelo serviço **caracterizar** (*text mining*). Deste modo, sempre que o

utilizador pretende obter informação sobre um determinado assunto, introduz um conjunto de palavras-chave, sendo devolvida uma tabela listando todos os documentos que contém alguma das palavras-chave.

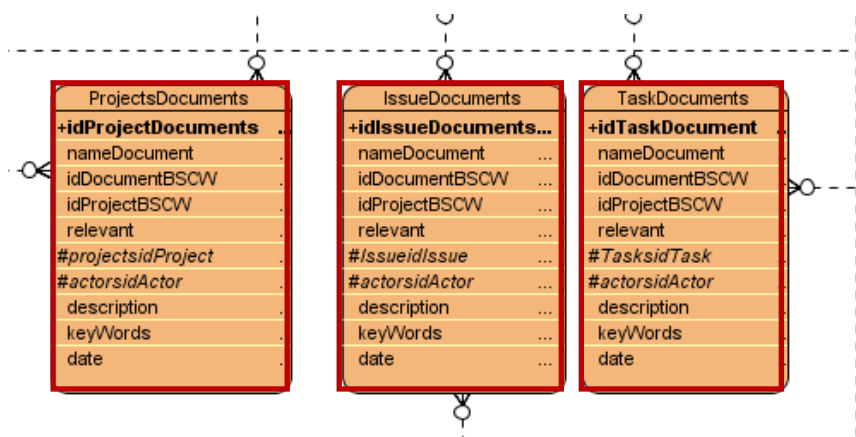


Figura 4.26: Excerto do DER (Ligações dos documentos)

4.2.2.4 Correspondência entre serviços *Miner*, métodos e tabelas

Todas as classes do *Miner* contêm funções específicas que permitem aplicar serviços e apresentar os resultados obtidos. De modo a mostrar as correspondências entre todos os serviços, as classes e as respectivas funções, assim como as tabelas da base de dados utilizadas, apresenta-se a tabela 4.2.

Tabela 4.2: Resumo dos serviços *Miner*

	Serviço	Descrição	Métodos	Tabelas
Reuniões	Caracterizar, classificar e obter informações para questões	Pesquisar questões semelhantes à questão em análise.	listSimilarIssueFromMeetings() similarIssueFromMeetings(...) parametersSimilarIssueMeetings()	Questões
Questões	Caracterizar, classificar e obter informações para questões	Pesquisar questões semelhantes	listSimilarIssue() similarIssue(...) parametersSimilarIssue()	Questões

Projectos	Prever	Estima o custo e tempo de execução do projecto	listPredictionProject() predictionProject(...) parametersPredictionProject()	Projectos e família de projectos
	Seleccionar actores	Elabora um <i>ranking</i> de participações de actores em projectos	listRankingActorsProject() rankingActorsProject() parametersActorsProject()	Tarefas e actores
	Caracterizar e classificar	Pesquisar informação sobre projectos semelhantes	listSimilarProjects() similarProjects(...) parametersSimilarProjects()	Projectos, família de projectos, extra projectos
Tarefas	Prever	Estima o custo e tempo de execução da tarefa	listPredictionTask() predictionTask() parametersPredictionTask()	Tarefas
	Seleccionar actores	Elabora um <i>ranking</i> de participações dos actores	listRankingActorsTask() rankingActorsTask() parametersActorsTask()	Tarefas e actores
	Caracterizar e classificar	Pesquisar informação sobre tarefas semelhantes	listSimilarTasks() similarTasks(...) parametersSimilarTasks()	Tarefas
Documentos	Caracterizar (TM)	Pesquisar documentos num determinado contexto	listSimilarDocuments() similarDocuments(...) parametersSimilarDocuments()	Documentos de projectos, tarefas e questões

4.3 Descrição do contributo das ferramentas mais importantes na implementação do *Miner*

4.3.1 Utilização do BSCW

Neste trabalho, a interface CoSKS foi integrada com o BSCW, de modo a permitir o armazenamento de dados, sob a forma de documentos.

Recorrendo a um conjunto de funções da *Application Programming Interface* (API) do BSCW, é possível criar pastas e subpastas de modo a organizar o armazenamento de documentos. Neste trabalho, a organização dos documentos foi efectuada de acordo com o ilustrado na figura 4.27 e uma ilustração mais detalhada da organização dos documentos encontra-se no **Anexo B** deste documento. As pastas principais correspondem a famílias de projectos e dentro destas encontram-se subpastas que representam cada projecto. Por sua vez, dentro de cada subpasta projecto são armazenados os documentos criados ou utilizados durante o desenvolvimento do mesmo.

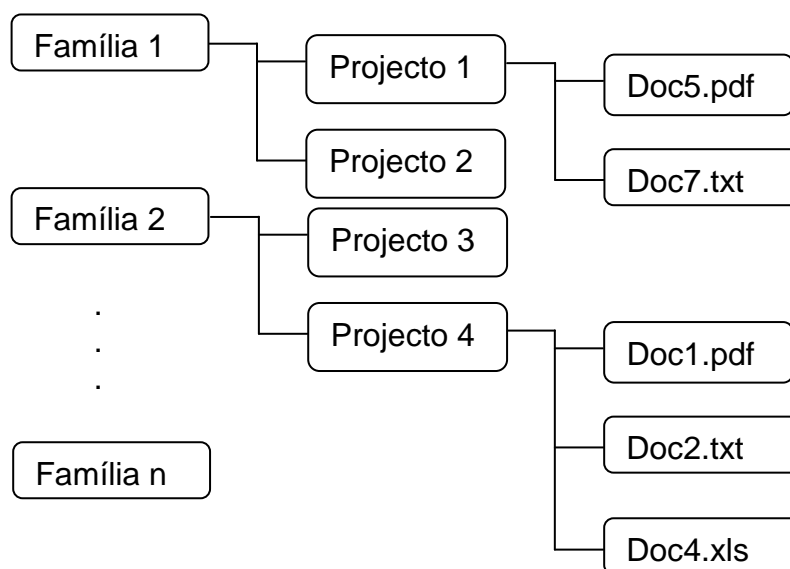


Figura 4.27: Estrutura do BSCW

Cada documento é caracterizado por um conjunto de atributos, que inclui o nome, a data assim como uma descrição. Deste modo quando há necessidade de adquirir informação para a realização de um serviço, o *Miner* aplica um algoritmo de TM no

atributo descrição de cada documento. Isto permite identificar um conjunto de documentos que são depois partilhados com o utilizador.

4.3.2 Utilização do RapidMiner

De acordo com a comparação realizada no capítulo 2 sobre ferramentas de DM, optou-se por utilizar o RapidMiner para aplicação das técnicas de DM neste trabalho. Inicialmente foram introduzidos dados de modo a tornar possível a aplicação de técnicas de DM pelo RapidMiner. Neste sentido criaram-se duas famílias de projectos, denominadas *carros* e *edifícios*, sendo em seguida criados 25 projectos distribuídos pelas duas famílias. Adicionaram-se participantes aos projectos, assim como questões e reuniões. Por sua vez, às questões associaram-se tarefas para as resolver e às tarefas participantes para as desempenhar. Aos vários projectos, questões e tarefas foram adicionados ficheiros que correspondem aos documentos consultados ou criados pelos participantes durante o desenvolvimento de projectos. Estabeleceu-se a ligação entre o RapidMiner e o MySQL de modo a importar os dados necessários e, para aplicação de serviços, procedeu-se de acordo com as seguintes etapas:

1. Realização de um SELECT à base de dados seleccionando os atributos necessários a cada serviço;
2. Escolha da técnica de DM a aplicar; e
3. Análise de resultados.

A título de exemplo, a figura 4.28 mostra o resultado da aplicação da técnica árvores de decisão sobre os atributos família, custo duração e criador de projecto. A árvore de decisão apresentada permite visualizar as relações entre as empresas responsáveis por projectos de acordo com a família, custo e duração dos mesmos.

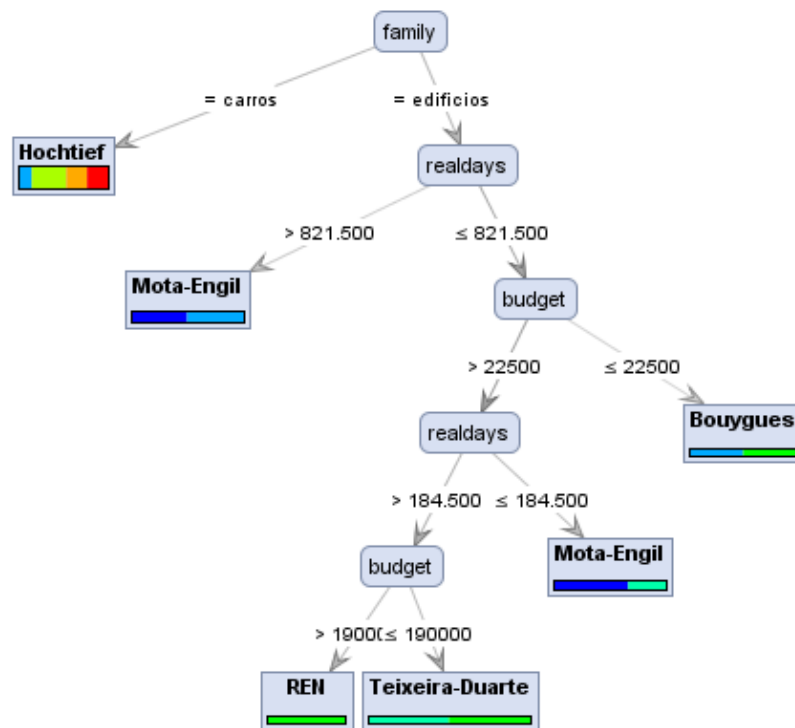


Figura 4.28: Árvore de decisão obtida no RapidMiner

O processo utilizado pelo RapidMiner para a aplicação de DM foi adoptado na implementação de serviços *Miner*. Assim, para cada serviço foram realizados os respectivos **SELECTs** à base de dados, de modo a obter os dados necessários para a sua aplicação. Na implementação dos serviços utilizaram-se árvores de decisão pois têm a vantagem de permitir expressar o conhecimento adquirido por intermédio de um conjunto de regras do tipo **IF-THEN-ELSE**, assim como permitem obter resultados intuitivos para qualquer utilizador. Por último traduziram-se os resultados obtidos na forma de tabelas para apresentação ao utilizador.

Para a realização de serviços que utilizam dados de texto, recorreu-se à técnica TM também disponibilizada pelo RapidMiner. Esta técnica permite realizar diversas pesquisas e acções em documentos de texto, nomeadamente contar o número de palavras, eliminar palavras, verificar a presença de determinada palavra, etc.. Novamente, os processos utilizados pelo RapidMiner foram adoptados para realizar algoritmos de TM à base de dados. Assim, quando há necessidade de analisar atributos de texto, o utilizador insere um conjunto de **palavras-chave**, que vão

constituir um **vector** de palavras. Em seguida é avaliada a presença de cada uma das palavras do vector nos atributos de texto em análise. De referir que os serviços de TM foram implementados de tal modo que o vector das palavras-chave não considera palavras com menos de três letras, nem caracteres que não sejam letras.

4.4 Pequeno exemplo de validação do sistema

Nesta secção do trabalho descreve-se um pequeno cenário adoptado para validação dos serviços *Miner*.

A interface utilizada para testar e validar o funcionamento do *Miner* foi o CoSKS. Após efectuado o login (figura 4.29) é possível aceder à página inicial da interface (figura 4.30).



The image shows a web browser window displaying the CoSKS login page. The header features the 'cospaces' logo and the text 'CoSpaces Knowledge Support'. The main content area is titled 'CoSKS - Login Page' and includes a welcome message: 'Welcome to CoSKS. Please, fill out the login form with the credentials given by the site administrator or your project manager.' Below this, there is a login form with two input fields: 'Username' (containing 'Joao Pinto') and 'Password' (with masked characters). A 'Login' button is positioned below the password field.

Figura 4.29: Página de acesso à interface CoSKS

Na secção delimitada pelo rectângulo vermelho (figura 4.30) estão representadas algumas das acções que se podem realizar, nomeadamente, criar, actualizar e

consultar projectos, reuniões, tarefas, adicionar actores a projectos ou tarefas. Também nesta secção é possível seleccionar os serviços *Miner* pretendidos. Por outro lado, qualquer resultado dos serviços será mostrado na secção limitada a verde.



Figura 4.30: Página inicial da interface CoSKS

O processo adoptado para a realização de qualquer serviço é semelhante. Consoante o serviço, é realizado um **SELECT** a uma ou mais tabelas da base de dados de forma a obter os dados para a execução do mesmo. Cada serviço é representado por um conjunto de **IF-THEN-ELSE** que representam o modelo de árvore de decisão adoptado e, consoante o valor dos parâmetros introduzidos, percorre-se a árvore até encontrar um resultado, o qual é mostrado na zona delimitada a verde da figura 4.30.

No serviço **prever** aplicado ao estudo de projectos, que se encontra-se ilustrado na figura 4.31 é possível oferecer estimativas do custo e duração dos projectos, sendo que a informação obtida será mais relevante se for aplicada a uma família específica de projectos.

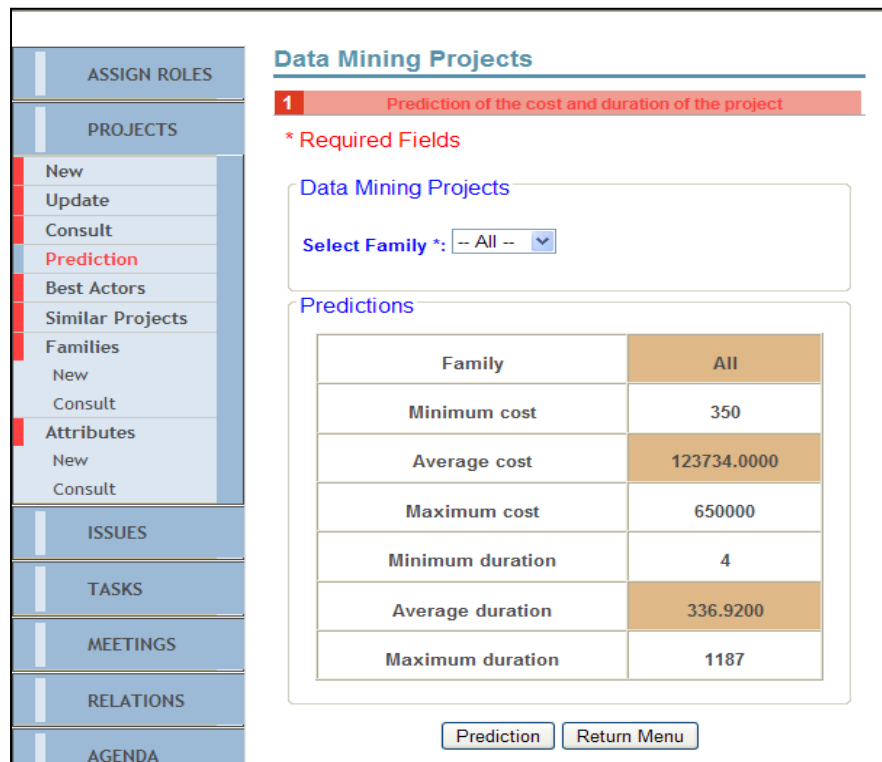


Figura 4.31: Resultados da utilização do serviço prever

Quando este serviço é aplicado a uma família de projectos específica, o *Miner* realiza um **SELECT** à base de dados para obtenção dos valores dos custos e tempos de execução de todos os projectos dessa família e determina o custo e tempo de execução médios. Por outro lado, se a opção for a estimativa desses mesmos valores, mas para todos os projectos presentes na base de dados, é realizado um **SELECT** aos valores dos custos e tempos de execução de todos os projectos e efectuado o mesmo cálculo (figura 4.31).

A figura 4.32 mostra os resultados obtidos pela utilização do serviço **seleccionar actores**. De modo a executar este serviço, é realizado um **SELECT** à tabela *actorsProjects*, a qual representa todas as participações de todos os actores nos

projectos. Com os dados em sua posse, o *Miner* calcula a percentagem de participação de cada actor em projectos.

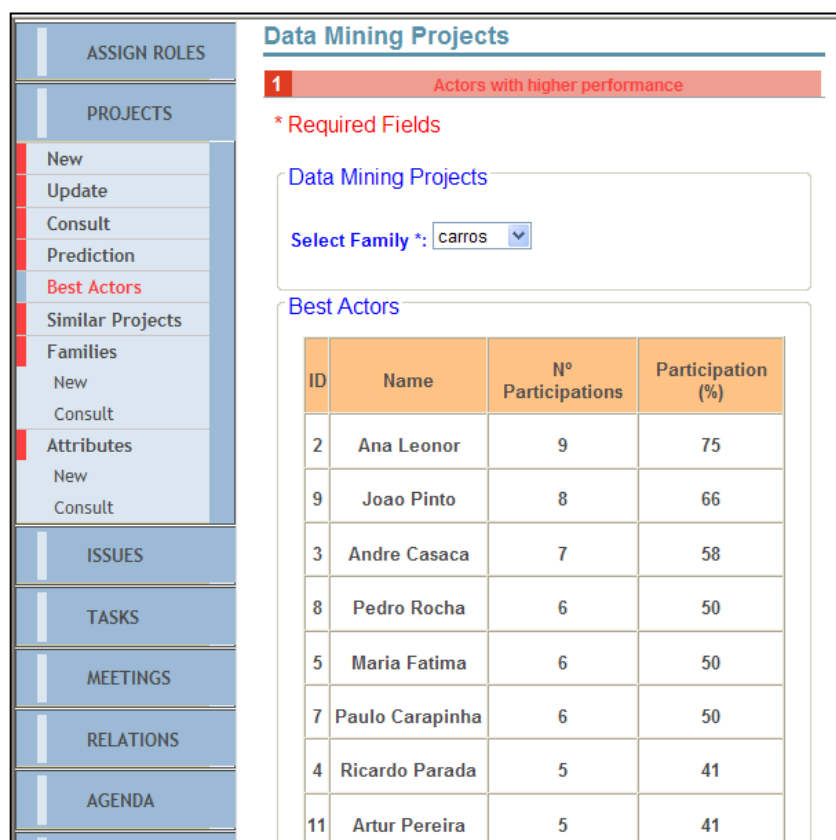


Figura 4.32: Resultados da utilização do serviço seleccionar actores

Este serviço permite a obtenção de uma lista ordenada dos participantes, de acordo com a frequência das suas participações em projectos de uma determinada família de projectos, neste caso, a família *carros*. Assim, a consulta desta lista pode facilitar a escolha de actores mais adequados para um novo projecto. No entanto, esta lista é organizada de acordo com o número de participações de cada actor, tratando-se por isso de uma avaliação quantitativa.

Na figura 4.33 exibem-se os resultados da aplicação dos serviços **caracterizar** e **classificar** que permitem obter uma lista de projectos incluindo determinadas características pretendidas. Estas características são discriminadas na forma de palavras-chave previamente introduzidas. Para o processamento destes serviços selecciona-se a opção *Similar Projects*, tal como mostrado na figura 4.33 e o *Miner*

realiza um **SELECT** à tabela *projects* de forma a obter todas as descrições de projectos. Em seguida, identificam-se as descrições que contêm as palavras-chave introduzidas. Deste modo, os projectos que incluem, nas suas descrições, pelo menos uma destas palavras-chave são seleccionados como projectos semelhantes. Os resultados aqui obtidos podem ser tornados mais específicos, uma vez que há a possibilidade de restringir a pesquisa a uma família particular de projectos.

The screenshot shows a web application interface for 'Data Mining Projects'. On the left is a vertical navigation menu with options: New, Update, Consult, Prediction, Best Actors, Similar Projects (highlighted in red), Families, New, Consult, Attributes, New, Consult, ISSUES, TASKS, MEETINGS, RELATIONS, AGENDA, MINUTES, TEMPLATES, and KNOWLEDGE. The main content area is titled 'Data Mining Projects' and contains a 'Select Family *' dropdown menu set to 'carros'. Below it is a 'Keywords *' text input field containing the text 'construção, repara edificios, ll ty + co'. Underneath is a section titled 'Similar Projects' which displays a table of results. At the bottom of this section are two buttons: 'Similar Projects' and 'Return Menu'.

ID	Titles	Descriptions
3	carro 1	construção de novo carro
7	carro 2	construção de novo carro
10	carro 3	construção de novo carro
14	carro 4	reparação do carro
19	carro 7	reparação do carro
21	carro 9	reparação do carro

Figura 4.33: Resultados da utilização dos serviços caracterizar e classificar

Alguns dos serviços descritos para o estudo de projectos, podem também ser utilizados para o estudo de tarefas, reuniões ou questões. Por exemplo, o serviço **seleccionar actores** pode ser aplicado para encontrar os actores mais apropriados para desempenharem determinadas tarefas. De igual modo, os serviços **caracterizar** e **classificar** podem ser utilizados para encontrar tarefas, questões ou reuniões, que se enquadrem num determinado contexto, descrito por intermédio de

palavras-chave. Outros exemplos da utilização destes serviços encontram-se no **Anexo C** deste documento.

5. Conclusões

Neste capítulo apresenta-se uma síntese do trabalho desenvolvido, com descrição dos objectivos alcançados e problemas encontrados. Por outro lado, é realizada uma análise do contributo deste trabalho discutindo possíveis alterações que venham a permitir obter melhores resultados.

5.1 Síntese do trabalho

O presente trabalho teve como objectivo principal o desenvolvimento de uma infra-estrutura de software, o *Miner*, baseada em conceitos de DM, que torne possível a obtenção de informação a partir do histórico de projectos armazenados numa base de dados.

No sentido de atingir o principal objectivo deste trabalho e tendo em conta que um dos requisitos essenciais para o funcionamento do *Miner* é a existência de dados sobre os quais são aplicados um conjunto de serviços, foi inicialmente estruturada uma base de dados, para permitir o armazenamento de dados relativos a projectos, nomeadamente actores, tarefas, questões, reuniões e documentos. Também se estabeleceu a integração do CoSKS com a tecnologia utilizada para armazenar os documentos utilizados pelos participantes de projectos durante o desenvolvimento dos mesmos.

O *Miner* foi implementado no seio de um ambiente colaborativo de apoio ao desenvolvimento de projectos que se baseia em dois conceitos essenciais: questões/ problemas e decisões. Deste modo, durante o desenvolvimento de um projecto são criados vários portais de decisão onde são abordados tais conceitos. Nos portais de decisão o *Miner* pode fornecer informação sobre casos históricos de projectos anteriores, facilitando a tomada de decisões e reduzindo o tempo dispendido na procura de boas opções em situações específicas.

Neste trabalho, o funcionamento do *Miner* permite fornecer informações às equipas de projectos, no sentido de permitir uma melhor compreensão dos problemas comuns e de indicar estratégias para superá-los, nomeadamente:

- Oferta de um conjunto de soluções que auxilie os actores de um projecto na resolução de um problema ou tarefa que possa surgir. Este conjunto de soluções é apresentado na forma de uma lista de documentos que incluem palavras-chave descrevendo o problema ou tarefa em questão;
- Identificação dos actores mais adequados para um determinado projecto ou tarefa, com a elaboração de um *ranking* baseado na frequência de participações em projectos ou tarefas semelhantes;
- Informação dos custos e duração médios da execução de projectos ou tarefas necessárias para resolver determinado problema. No caso específico das tarefas, esta informação possibilita que o *Companion* informe o gestor do impacto, a nível de custos ou duração, que a realização de uma nova tarefa pode vir a ter no projecto em desenvolvimento; e
- Apresentação de um conjunto de questões que se relacionem com as questões abordadas durante uma reunião. A pesquisa é realizada com base num conjunto de palavras-chave que caracterizem a questão em análise. Uma vez identificadas as questões semelhantes, o gestor do projecto pode consultar cada uma e identificar tarefas realizadas para a sua resolução, assim como documentos que foram consultados, etc..

5.2 Contribuição da pesquisa

O trabalho aqui descrito contribui para melhorar o processo de colaboração entre indivíduos, no contexto do projecto CoSpaces.

A construção de uma base de dados estruturada de modo a permitir o armazenamento de dados relativos ao desenvolvimento de projectos, assim como a integração do CoSKS com a tecnologia BSCW constituem uma mais-valia no seio de um ambiente colaborativo pois possibilitam a partilha de dados entre indivíduos dispersos geograficamente.

Também a implementação do *Miner*, com a produção de informação a partir dos dados armazenados, contribui para melhorar os processos de tomadas de decisão em grupo e estimular a interacção entre os indivíduos.

Parte do trabalho realizado no âmbito desta dissertação encontra-se descrito na seguinte publicação:

- Costa, R., Lima, C., Antunes, J., Figueiras, P., & Parada, V. (2010). Knowledge Management Capabilities Supporting Collaborative Working Environments in a Project Oriented Context. *2nd European Conference on Intellectual Capital*, (pp. 1-9). ISCTE Lisbon University Institute, Lisbon, Portugal and Polytechnic Institute of Leiria, Portugal.

5.3 Trabalhos futuros

O *Miner* desenvolvido permite utilizar um conjunto de serviços de modo a extrair informação a partir de dados armazenados. Contudo, de modo a tirar maior proveito dos serviços referidos, estes devem ser testados futuramente num ambiente real.

Para além do ambiente real, a interacção com indivíduos experientes em várias áreas de projectos, poderá vir a permitir o melhoramento dos serviços oferecidos pelo *Miner*. A experiência destes indivíduos pode facilitar a selecção dos dados mais indicados para aplicação de cada serviço, assim como melhorar a estruturação da base de dados, de modo a que todos os atributos importantes sejam contemplados.

Referências Bibliográficas

- [1] Drucker, P. (1964). *Knowledge worker: new target for management*. Christian Science Monitor.
- [2] Nonaka, I. & Takeuchi, H. (1995). *The knowledge-creating company*. New York: Oxford University Press.
- [3] Grey, D. (1996). *What is knowledge management?* The Knowledge Management Forum.
- [4] Frapaolo, C. (2002). *Knowledge Management*. Capstone Publishing, John Wiley.
- [5] Information Week (2003). *Ten Principles for knowledge Management Success*.
- [6] Kazi, A.S., Puttonen, J., Sulkusalmi, M., Välikangas, P. and Hannus, M. (2001) Knowledge Creation and Management: the Case of Fortum Engineering Ltd., Knowledge Management in the Sociotechnical World: the Graffiti Continues, SpringerVerlag, pp.153-169.
- [7] Davenport, T. H. & Prusak, L. (1998). *Working Knowledge: How Organizations Manage What They Know*. Harvard Business School Press, Boston.
- [8] Dalkir, K. (2006). *Knowledge Management in Theory and Practice*.
- [9] Wiig, K. M. (1993). *Knowledge Management Foundations: Thinking About Thinking – How People and Organizations Represent, Create and Knowledge*. Aslington: Schema.

[10] Savage, C (1996). *Fifth Generation Management*. Boston: Butterworth-Heinemann.

[11] Polanyi, M. (1966). *The tacit dimension*. Gloucester, MA: Peter Smith.

[12] Nonaka, I. (1991). *The knowledge-creating company* Harvard Business Review, Boston: Harvard University Press, 96-104.

[13] Collis, D. & Montgomery, C. (1995). *Competing on resources: strategy in the 1990s*. Harvard Business Review, 118-129.

[14] Leonard-Barton, D. (1995). *Wellsprings of Knowledge*. Harvard Business School Press, Boston.

[15] Petrash, G. (1996). *Dow's journey to a knowledge value management culture*. European Management Journal 14 (4), 365-373.

[16] Holsapple, C. W. & Joshi, K. D. (2001). *Organizational knowledge resources*. Decision Support Systems, 39-54.

[17] Skryme, D. (1997). *Knowledge Management*.

[18] Skryne, D. & Amidon, D. (1997). *Creating the Knowledge-Based Business: key lessons from an internacional study of best practice*. Business Intelligence, London.

[19] Ellioy, S. & O'Dell, C. (1999). *Sharing knowledge & best practices: The hows and whys of tapping your organization's hidden reservoirs of knowledge*. Health Forum Journal 42 (3), 34-37.

[20] Pinto, A. (2006). *Os três pilares na gestão de conhecimento*. Revista espaço acadêmico nº58.

- [21] Simard, A. (2006). *Knowledge Markets: More than Providers and Users. Knowledge Creation Diffusion Utilization.*
- [22] Accorsi, F. L. & Costa, J. P. (2008). *A interação apoiada pela Gestão do Conhecimento e pelos Serviços Colaborativos.* Revista TEXTOS de la CiberSociedad, 13. Temática Variada.
- [23] Alwert, K. & Hoffmann, I. (2003). *Knowledge Management Tools.* In Mertins K., Heisig P. & Vorbeck J. (Ed.), *Knowledge Management: Concepts and Best Practices.* Heidelberg: Springer-Verlag, 114-150.
- [24] Tyndale, P. (2002). *A taxonomy of knowledge management software tools: origins and applications.* *Evaluation and Program Planning.*
- [25] Ngai, E. W. T. & Chan, E. W. C. (2005). *Evaluation of Knowledge Management Tools Using AHP.* *Expert Systems with Applications*, 1-11.
- [26] Bafoutsou, G. & Mentzas, G. (2002). *Review and functional classification of collaborative systems.* *International Journal of Information Management*, 22, 281-305.
- [27] Long D. & Fahey L. (2000). *Diagnosing cultural barriers to knowledge management.* *Academy of Management Executive*, 14, 4, 113-127.
- [28] Fayyad, U. M. (1996). *Data mining and knowledge discovery: making sense out of data.* *IEEE Expert*, 20-25.
- [29] Goebel, M. & Gruenwald, L. (1999). *A survey of data mining and knowledge discovery software tools.* *SIGKDD Explorations*, 1, 20-33.

- [30] Lee, J., Hsueh, S., & Tseng, H. (2008). *Utilizing data mining to discover knowledge in construction enterprise performance records*. *Journal of Civil Engineering and Management*, 14(2), 79-84.
- [31] Banks, D. L. (2010). *Statistical data mining*. John Wiley & Sons, Inc. WIREs Comp Stat, 2, 9-25.
- [32] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. & Verkamo, A. (1996) *Fast discovery of association rules*. In Advances in KDD. MIT Press.
- [33] Freitas, J. (2006). *Uso de técnicas de Data Mining para Análise de Bases de Dados hospitalares com Finalidades de Gestão*. Phd thesis. Faculdade de Economia da Universidade do Porto.
- [34] Luan, J. (2002). *Data Mining and Knowledge Management in Higher Education - potential applications*. Presentation at AIR Forum, Toronto, Canada.
- [35] Hearst, M. (2003). *What is Text Mining?*
- [36] Grobelnik, M. & Mladenic, D. (2005). *Text Mining and Web Mining*, presented at ACAI-05 Advanced Course on Knowledge Discovery, "Jožef Stefan" Institute Ljubljana, Slovenia.
- [37] Liddy, E.D. (2000). *Text Mining*. Bulletin of American Society for Information Science & Technology.
- [38] Cios K., Swiniarski R., Pedrycz W. & Kurgan L. (2007). *Data Mining - A Knowledge Discovery Approach*. XV, 606p., Harcover.

- [39] Morik, K. & Scholz, M. (2004). *The MiningMart Approach to Knowledge Discovery in Databases*. Intelligent Technologies for Information Analysis, N. Zhong and J. Liu (eds), Springer, 47–65.
- [40] Demšar, J. & Zupan, B. *Orange: From experimental machine learning to interactive data mining*, White Paper (<http://www.ailab.si/orange>), Faculty of Computer and Information Science, University of Ljubljana.
- [41] Rakotomalala, R. (2005). *TANAGRA: un logiciel gratuit pour l'enseignement et la recherche*. Proc. of the 5th Journ'ees d'Extraction et Gestion des Connaissances 2, 697–702.
- [42] Witten, I.H. & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Second Edition (Morgan Kaufmann, San Francisco) 525.
- [43] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M. & Euler, T. (2006) *YALE: Rapid Prototyping for Complex Data Mining Tasks*. Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1–6.
- [44] Santos, M. F. e Azevedo, C. (2005). *Data Mining, Descoberta de Conhecimento em Bases de Dados*, FCA – Editora de Informática.
- [45] King, D. (2004). *Numerical Machine Learning*, Technical Report CS 4803B, Georgia Technology College of Computing.
- [46] Alcalá-Fdez, J., Sánchez, L. García¹, S., del Jesus, M. J., Ventura, S., Garrell, J. M., Otero, J., Romero, C., Bacardit, J., Rivas, V. M., Fernández, J. C. & Herrera, F. (2005). *KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems*. *Knowledge Creation Diffusion Utilization*.

[47] Costa, R., Lima, C., Antunes, J., Figueiras, P., & Parada, V. (2010). *Knowledge Management Capabilities Supporting Collaborative Working Environments in a Project Oriented Context. 2nd European Conference on Intellectual Capital*, (pp. 1-9). ISCTE Lisbon University Institute, Lisbon, Portugal and Polytechnic Institute of Leiria, Portugal.

[48] Lima, C., Costa, R., Maló, P., & Antunes, J. (2010). *A Knowledge-based approach to support decision making process in project-oriented collaboration*. To appear in 11th European Conference on Knowledge Management. Vila Nova de Famalicão, Portugal.

[49] Netbeans IDE 6.8, <http://netbeans.org/community/releases/68/>.

[50] JavaServer Pages Technology, <http://java.sun.com/products/jsp/>.

[51] Relational Persistence for Java and .NET, <http://www.hibernate.org/>.

[52] Apache Tomcat, <http://tomcat.apache.org/>.

[53] Visual Paradigm, <http://www.visual-paradigm.com/>.

[54] MySQL, <http://www.mysql.com>.

[55] Struts 2, <http://struts.apache.org/2.1.8.1/index.html>.

[57] BSCW, <http://www.bscw.de/english/index.html>

.

Anexo A - Diagrama de Entidade Relação

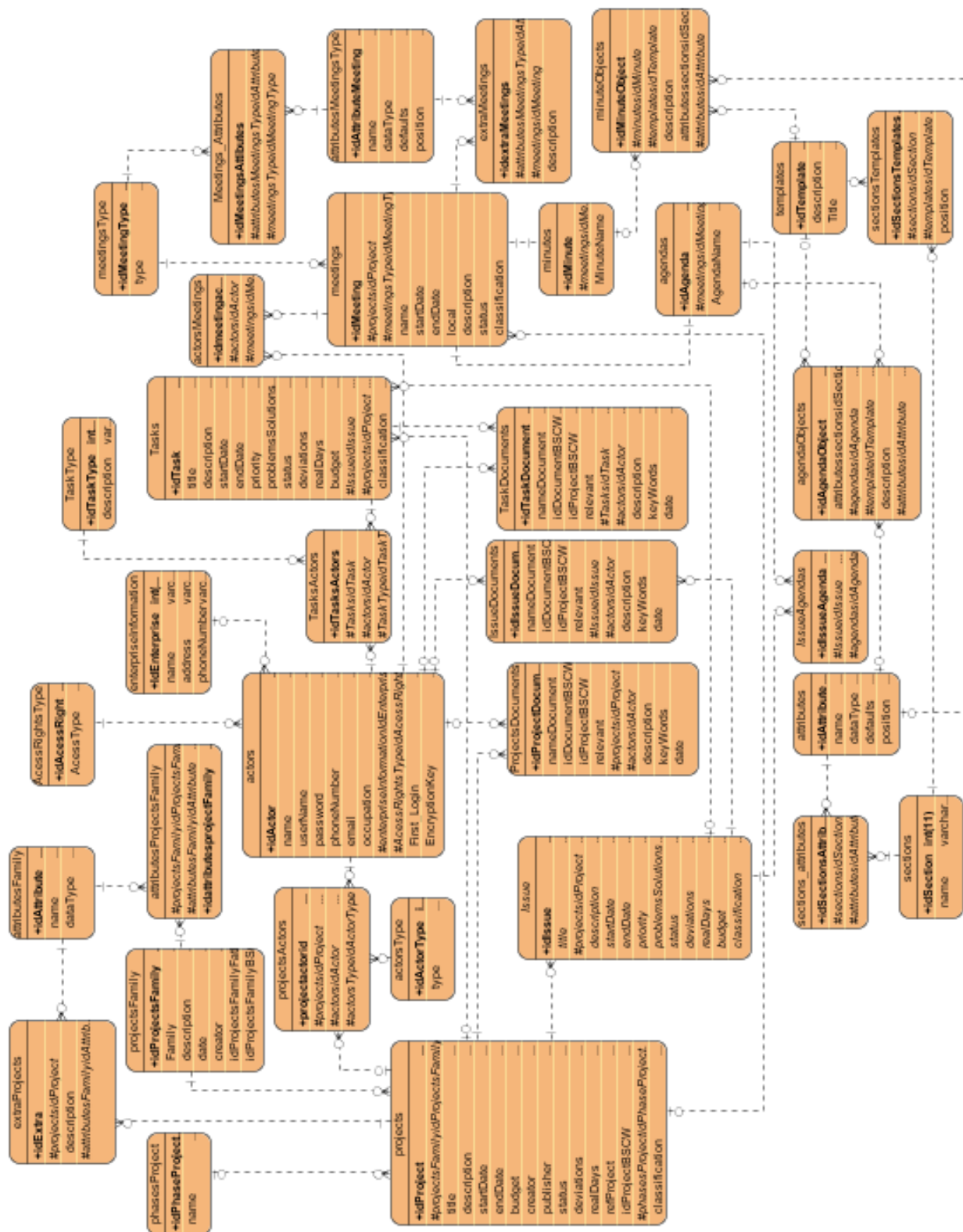


Figura A.1: DER

Anexo B - BSCW

The screenshot displays the BSCW web interface. At the top, there is a navigation bar with the BSCW logo, a 'Logout' button, and a menu with options: Arquivo, Editar, Exibir, Opções, Ir para, and Ajuda. Below the menu is a toolbar with various icons for file operations. The main area shows a workspace for 'jna' with a search bar labeled 'Buscar'. A table lists the contents of the workspace, including folders and documents. A tooltip is visible over the 'Document2.txt' entry, showing its metadata.

Ícone	Nome	Ação	Size	Creator	Proprietário	Prior Data	Eventos
Folder	carros creator: Bouygues	▼	13	jna	jna	2010-07-15 13:13	🔔
Folder	edificios creator: Mota-Engil	▼	13	jna	jna	2010-07-06 18:29	🔔
Folder	edificio 1 budget: 500000	▼	4	jna	jna	2010-07-12 12:55	🔔
Document	Document 37.txt Actor: Vitor Parada	▼	26.5 K	jna	jna	→ 2010-07-12 12:55	🔔
Document	Document1.txt Actor: Vitor Parada	▼	26.5 K	jna	jna	→ 2010-07-12 11:36	🔔
Document	Document2.txt Actor: Vitor Parada	▼	26.5 K	jna	jna	→ 2010-07-12 02:27	🔔
Document	Document3.txt Date: Mon Jul 12 00:00:00 BST 2010 Description: manutenção de carro Actor: Vitor Parada	▼	26.5 K	jna	jna	→ 2010-07-12 12:54	🔔
Folder	edificio 10	▼	0	jna	jna	2010-07-06 17:46	🔔

Figura B.1: BSCW

Anexo C - Exemplos da utilização do *Miner*

Na figura C.1 apresentam-se os resultados da utilização do serviço **obter informações para questões**. Após a selecção do *Similar Issue* da figura C.1, é realizado um **SELECT** a todas as descrições das questões relacionadas com a família de projectos edifícios. Em seguida, identificam-se as descrições que contêm as palavras-chave introduzidas. Deste modo, as questões que incluem, nas suas descrições, pelo menos uma destas palavras-chave são seleccionadas como questões semelhantes.

Data Mining Issue

1 Similar Issue

* Required Fields

Data Mining Issue

Select Family *: edifícios

keywords *: infiltraç

Similar Issues

ID	Titles	Descriptions
2	edificio 13 issue 2	infiltração nas cozinhas
6	Problema tipo 5	canalização deficiente, infiltrações de água nas casas de banho..

Similar Issue Return Menu

Figura C.1: Resultados do serviço obter informações na análise de questões

Os resultados obtidos podem ser úteis no decorrer de uma reunião por permitirem encontrar informação sobre questões semelhantes à que se encontra em debate.

O processo adoptado para as questões pode ser utilizado na pesquisa de tarefas e projectos semelhantes.

A figura C.2 ilustra os resultados obtidos após a aplicação do serviço **classificar (text mining)**, como objectivo de encontrar documentos contendo determinadas palavras-chave. Note-se que a realização de TM obedece a algumas regras nomeadamente, as palavras-chave tem de possuir três ou mais letras, não são considerados algarismos nem outros caracteres específicos assim como palavras repetidas.

Data Mining Documents

Keywords *: motor, construir, montar,

Similar Documents

ID	Name document	Description
3	MontagemReparação.pdf	reparar e montar motor de carros
2	Document27.txt	apontamentos sobre como construir uma vivenda junto à praia.

Similar Documents Return Menu

Figura C.2: Resultados da utilização do serviço classificar (TM)

